12-10-2021

A longitudinal analysis of pathways to computing careers: Defining broadening participation in computing (BPC) success with a rearview lens

Mercy Jaiyeola
Mississippi State University, darajay2008@gmail.com

A longitudinal analysis of pathways to computing careers:
Defining broadening participation in computing (BPC)
success with a rearview lens

By

Mercy Jaiyeola

Approved by:

Shahram Rahimi (Major Professor)
Sarah B. Lee (Minor Professor)
Allen S. Parrish
Mike Taquino
T.J. Jankun-Kelly (Graduate Coordinator)
Jason M. Keith (Dean, Bagley College of Engineering)

A Dissertation
Submitted to the Bagley College of Engineering
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Computer Science
in the Department of Computer Science and Engineering

Mississippi State, Mississippi
December 2021

Name: Mercy Jaiyeola

Date of Degree: December 10, 2021

Institution: Mississippi State University

Major Field: Computer Science

Major Professors: Shahram Rahimi, Sarah B. Lee

Title of Study: A longitudinal analysis of pathways to computing careers: Defining broadening participation in computing (BPC) success with a rearview lens

Pages in Study: 123

Candidate for Degree of Doctor of Philosophy

Efforts to increase the participation of groups historically underrepresented in computing studies, and in the computing workforce, are well documented. It is a national effort with funding from a variety of sources being allocated to research in broadening participation in computing (BPC). Many of the BPC efforts are funded by the National Science Foundation (NSF) but as existing literature shows, the growth in representation of traditionally underrepresented minorities and women is not commensurate to the efforts and resources that have been directed toward this aim.

Instead of attempting to tackle the barriers to increasing representation, this dissertation research tackles the underrepresentation problem by identifying what has worked (leveraging existing real-world data) to increase representation. This work studies the educational pathways of persons who have successfully transitioned into the computing workforce and identifies the common roadmaps that have contributed to retention, persistence, and success in attaining computing employment. Descriptive statistics, Logistic regression, Classification algorithms, Clustering, and Predictive analytics were employed, using the Stata statistical tool and Orange

Data Mining tool on real-world data, to identify educational pathways that have resulted in successful employment outcomes for women and blacks in computing.

The results of this analysis have highlighted key information that is capable of informing future “Broadening Participation in Computing” (BPC) efforts. This is because the information will enable researchers and decision makers to have a clearer picture of what educational choices have resulted in favorable outcomes for underrepresented minorities and women in computing; and consequently, researchers and decision makers would be able to more accurately target their BPC efforts to achieve optimal results. This knowledge can also be applied in career advising for young students who are trying to chart their path into computing, providing insight into alternative pathways.

DEDICATION

I want to dedicate this dissertation to the Almighty God: my source of life, intellect and inspiration.

ACKNOWLEDGEMENTS

I want to appreciate my family for being there for me always. They have encouraged me, prayed for me, supported me, helped me, sponsored me, and advised me. A big thank you to my Dad – Engineer Bimbo Jaiyeola, my Mum – Mrs. Yemi Jaiyeola, my sister – Grace Jaiyeola, and my brother – David Jaiyeola.

I also want to thank my ever-supportive colleagues and friends: Abiodun Ajayi, Theophilus Okunlola, Temitope Emoruwa, Damiloju Aroyewun, Rachel Agboola, Damilola Aroyewun, Feyi Oyeniyi, Tobi Lasisi, Delma Nieves-Rivera, Sayantani Dutta, Marcus Brumfield, Halleluyah Aworinde, Roger Garcia, among others, for helping me, supporting me, praying for me, and encouraging me throughout my program and during my dissertation.

Third, I want to thank my advisors and committee members for being my sounding board, my teachers and mentors, and for sacrificing their time and expertise to help me throughout this journey. Big thanks to Dr. Sarah Lee, Dr. Allen Parrish, Dr. Shahram Rahimi, Dr. Mike Taquino, and Dr. Ed Swan. I have learnt so much from you; thank you.

Thanks to Dr. Kim Stowers of the University of Alabama and thanks to Dr. Steve Grice and Dr. Shane Warren of NSPARC, Mississippi State University for helping me through my research and data collection process.

I also want to express my gratitude to Dr. Maxwell Young, my first advisor at Mississippi State University for teaching me how to be a great researcher.

Finally, I want to acknowledge my own tenacity, can-do spirit, and I appreciate myself for not giving up when the going seemed tough.

Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Introduction

There is an abundance of open computing jobs in the United States of America (USA), but the rate of supply of computing professionals to the computing workforce has lagged behind the level of demand for these talents (McClelland, 2001; Peckham et al., 2007; Doerschuk et al., 2009). In this research, a number of gaps have been identified as sources of disruption along the path into the computing workforce. The gaps are the skills gap, gender gap, and the racial gap.

The skills gap occurs when there are not enough qualified professionals to fill up available positions in the computing industry (Evans, 2017). The skills gap could likely be a reflection of the quality of instruction or education received prior to entering the job market. The gender gap describes a disparity in the gender composition of the computing field and workforce (McClelland, 2001; Lynn et al., 2003; Stout & Camp, 2014). When the distribution of professionals in the workforce is skewed along gender lines, there is said to be a gender gap. Similarly, the racial gap depicts an uneven distribution among racial groups within the computing workforce (McClelland, 2001).

These employment gaps have been widely studied by researchers in the computing and social sciences fields; and in order to increase the supply of computing professionals to the workforce, it is important to devise effective means of closing the skills, gender, and racial gaps in the computing field (McClelland, 2001).

This dissertation research studies these gaps, particularly along the racial and gender dimensions, and seeks a different perspective for understanding the underrepresentation problem in computing than what has been seen in the existing literature. This research will illustrate a roadmap of how current computing professionals have reached their employment outcomes and will map this knowledge to distinguishing characteristics of those professionals such as gender, race, and educational preparation. This will inform work in broadening participation in computing (BPC), providing an evidence-based alternative view of the interventions and educational pathways that are representative of persons historically underrepresented persisting into computing jobs, such as women and racial minorities, navigating and persisting in those roles.

1.2 Purpose of Study and Research Questions

Efforts to increase the participation of groups historically underrepresented in computing studies, and in the computing workforce, are well documented. It is a national effort with funding from a variety of sources being allocated to research in broadening participation in computing (BPC). Many of the BPC efforts are funded by the National Science Foundation (NSF) (James & Singer, 2016; Hoffman et al., 2019; Bruckman et al., 2009; Doerschuk et al., 2009) but as existing literature shows, the growth in representation of traditionally underrepresented minorities and women is not commensurate to the efforts and resources that have been directed toward this aim (Varma, 2018; Funk & Parker, 2018).

Thus, this research proposes to bring a different perspective to understanding and addressing the underrepresentation problem in computing. Many of the BPC efforts that are evident in the literature have identified one or more barriers that have impeded the representation of the underrepresented minorities and have attempted to tackle these individual barriers. Instead of attempting to tackle the barriers (what does not work), this dissertation research will tackle the underrepresentation problem by identifying what has worked (leveraging existing real-world data) to increase representation. This work studies the educational pathways of persons who have successfully transitioned into the computing workforce and will identify the common roadmaps that have contributed to retention, persistence, and success in attaining computing employment. This strategy promises to be impactful because the identification of factors that increase representation across racial and gender dimensions will inform the direction of future investments in BPC efforts.

In pursuit of this goal, the following research questions will be addressed:

1. What are the common themes across educational pathway experiences that emerge from the analysis of computing professionals' data across racial and gender dimensions?
2. Which of these common experiences result in successful long-term (greater than 3 years) employment outcomes in the technology sector for women and blacks?
3. How do the findings of this study inform national investment in broadening participation efforts that seek to increase racial and gender diversity in the computing workforce?

1.3 Scope/Location of Study

Conducted from Mississippi State University, Starkville, USA, this research studies the computing education and workforce landscape within the United States of America with specific focus on the outcomes of computing studies and employment in the Southern region of the USA.

1.4 Broader Impacts

Answering the research questions (stated earlier) would enable researchers and decision makers to have a clearer picture of what educational choices have resulted in favorable outcomes for underrepresented minorities and women in computing; and consequently, researchers and decision makers would be able to more accurately target their BPC efforts to achieve optimal results.

If the data shows that certain educational choices result in successful employment outcomes for specific groups, the research community can recommend these educational choices for efforts designed to increase diverse representation in the computing workforce. This knowledge can also be applied in career advising for young students who are trying to chart their path into computing, providing insight into alternative pathways. However, it is important that academic and career advising is not confined to any one pathway based on characteristics of the target individual, regardless of the outcome of this work.

As national organizations continue to fund efforts to broaden participation, the results of these research would inform the direction of future investment in computing education and BPC efforts, providing evidence on interventions and educational pursuits that have frequently resulted in positive computing career outcomes. Depending on the results of this study, there may be opportunity for further investigation to identify why some pathways are least successful. Perhaps those are where incremental BPC investments should occur, to remove apparent barriers in those channels.

1.5 Definitions of Common Terms

Educational Pathways: The series of educational choices made by individuals as they navigate into their career (National Science Foundation, 2011; MDRC, 2015).

Traditional Educational pathways: This refers to the nationally-accepted standard for formal education e.g., K-12, High school, college (Wikipedia, 2019).

Alternative educational pathways: This refers to less formal/standardized educational programs and qualifications e.g., after-school classes, Certifications, Coding bootcamps (Wikipedia, 2019).

Computing educational pipeline: The traditional educational pathway leading to a computing degree (Bruckman et. al., 2009; Fealing et. al., 2015; Adrion, et. al., 2008).

Representation (of gender and racial minorities): Being present or found in a larger context (Cambridge Dictionary, 2021).

Underrepresentation: Having a low representation, especially when compared to other groups (Merriam-Webster, n.d.).

Underrepresented minorities: Minority groups whose representation in a specific context is less than their representation in the entire population (Penn State, n.d.).

Computing field: The entire computing landscape; including the body of knowledge and the computing profession (Denning, 2007; Locsin, n.d.).

Computing workforce: The individuals that are engaged or employed in the computing profession and industry (Merriam-Webster, n.d.).

Broadening participation in computing: The act of increasing the involvement of historically underrepresented groups in the computing field (National Science Foundation, 2021).

Employment outcome: This is the state of entering into, progressing in, or persisting in full-time or part-time employment (New York State Education Department, 2018).

Skills gap: The gap between the skills of college graduates and the skills required by employers (Clear et. al., 2019).

Racial gap: The difference between employment outcomes across different racial groups (Ajilore, 2020; Williams & Wilson, 2019).

Gender employment gap: The difference between male and female employment outcomes (Eurofound, 2016).

1.6 Limitations

This research bases its findings on publicly available social media data (from LinkedIn) which has its own peculiar limitations. Here are some limitations that have been identified:

- This research largely attributes successful employment outcomes to the educational choices that participants made along their pathway to employment. This is a limitation because there are other variables (not shown in the LinkedIn data) that influence employment outcomes.
- Since the data is usually manually updated by the participants, there is no guarantee that the data employed in this research is up to date (as the user might have not updated it).
- Since the data for this research is retrieved from a professional career website, many participants would not disclose their unprofessional (less-than-ideal) employment outcomes; hence the full picture of their journeys might not be available for analysis.
- Certain assumptions might be made that might not reflect the true situation of the participants. For example: a participant who is unemployed after graduation might be labelled by this research as having an unsuccessful employment outcome whereas they might be out of the job market by choice and might be experiencing even more successful outcomes (according to their own personal definition of success).
- The sample size of the data is small. This is also a limitation.

CHAPTER II

LITERATURE REVIEW

2.1 BACKGROUND

2.1.1 Composition of the Computing Field

As alluded in the topic of this dissertation, the scope for the research is the computing field, which is a subset of the STEM (Science, Technology, Engineering, Mathematics) field. According to Computing Curricula 2005 (Shackelford et al., 2006), the computing field includes these major disciplines: Computer Engineering (CE), Computer Science (CS), Information Systems (IS), Information Technology (IT), and Software Engineering (SE). As seen in Computing Curricula 2020 (Clear et al., 2019), newer computing disciplines such as Artificial Intelligence and Data Science have been added to the list of disciplines within the computing field. The computing field is a major area of focus in this research because as seen in Exter et al. (2018), computing employment positions take longer to be filled than other types of professions. Also, in a news article (Carreon, 2019, para. 12) published by the Northeast Mississippi Daily Journal, a representative of a technology company said that in the state of Mississippi, “there are currently almost 1,000 unfilled job openings due to a shortage of qualified IT workers”.

2.1.2 The Computing Employment Landscape

For employees with computing college degrees, in order to accurately identify what might count as underemployment in the computing workforce or employment in non-computing jobs, we would need to investigate what is considered a computing job. According to Computing Curricula

2005 (Shackelford et al., 2006), here are some fields that computing graduates might develop competencies in: Algorithms, Application programs, Computer programming, Hardware and devices, Human-computer interface, Information systems, Information management (Databases), IT resource planning, Intelligent systems, Networking and communications, and Systems Development through Integration.

Similarly, several research papers have identified specific job categories, roles, titles or positions that computing graduates might hold, in addition to the skills required for such roles. In El-Agamy & Tsuda (2013), the author presented the Information Technology Skill Standards (ITSS), which was published by the Japanese government in order to specify what the IT job categories are, and what skills are needed for these job categories. These defined skills provide a template that can be used to design educational programs to produce graduates fit for these job categories. The IT job categories defined by Japan's ITSS are: Education, IT Service Management, Customer Service, Software Development, Application Specialist, IT Specialist, Project Management, IT Architect, Consultant, Sales, and Marketing.

According to Miller & Voas (2008), there are several IT job-titles (computer scientist, computer engineer, systems analyst, computer programmers, computer software engineers, computer support specialists, database administrators, network systems and data communications analysts, etc.) which were condensed by the authors into 3 major groups of jobs in computing: Computer scientist, software engineer, and IT professional. According to the authors, ideally, computer scientists study things about computers and how they work, software engineers build systems and solutions that make computers work better, while IT professionals apply knowledge about systems and software to meet users' needs (Miller & Voas, 2008).

Published in year 2000, a research paper (Job Categories and Education Requirements in Information Technology, 2000) identified 8 major job categories (and sample job titles) within the IT field of computing: Analysis and Integration of Business Systems (e.g. Systems Analyst), Development and Administration of Databases (e.g. Database Developers), Design and Administration of Networks (e.g. Network Administrators), Programming and Software Engineering (e.g. Programmer/Analyst), Technical Support Specialists (e.g. Technical Support Engineers), Development and Administration of Websites (e.g. Web Designers), Digital Media (e.g. Animators), and Technical Writing and Editing (e.g. Technical Writers).

2.1.3 Employability and the Concept of Successful Employment Outcomes

There have been several studies and discussions around the topic of the skill sets and employability of employees in various fields in the United States, and globally too. According to the Commission on Higher Education and Employability (2018, p. 11), “Employability is a set of achievements — skills, understandings and personal attributes — that make graduates more likely to gain employment and be successful in their chosen occupations, benefiting themselves, the workforce, the community and the economy”.

In the context of this research, the success of a person’s employment outcome is dependent on about 5 factors:

1. Whether they were able to gain computing employment: Getting employed in a computing job is seen as a successful employment outcome whereas getting employed in a non-

computing position is seen as an unsuccessful outcome, in the context of this research (Negoita & Dunham, 2013; Workforce Connections, 2015).

2. How soon they were able to gain their computing employment: The longer an individual is unemployed, the more difficult it will be to find employment (Indeed Editorial Team, 2021). Therefore, the faster an individual is able to gain computing employment, the more successful they are, in the context of this research. After college, it takes the average graduate three to six months to get a job (University of Washington, 2021). Within this research, a person who gets employed within a year of the completion of their educational program(s) is seen as successful whereas a person who gains their computing employment after a wait period of over one year is not as successful.
3. The rating of their company or employer: The higher a person's employer's rating is, the more successful their employment is assumed to be. For example, having work experience at a Fortune 500 company is usually perceived as a sign of success (Joyce, n.d.). A computing employment at a Fortune500 company is seen as more successful than one at a Non-Fortune500 company.
4. Their salary level: An employee's salary level is usually reflective of the type of position they are employed in (Negoita & Dunham, 2013; Workforce Connections, 2015; Brooke & Revell, 2009). Within this research, a person with a low annual income (less than 50,000 USD) and a person with a medium annual income (between 50,000 and 100,000 USD) are seen as less successful than a person with a high annual income (greater than 100,000 USD).
5. Whether they are able to persist in a computing field (Negoita & Dunham, 2013; Lindsay & Babb, 2015): Within this research, an employee that is able to remain in the computing

workforce for at least 3-5 years after they enter the workforce is seen as more successful than those who aren't able to persist for that long in those computing positions.

2.2 SKILLS GAPS IN COMPUTING EMPLOYMENT

2.2.1 Skills gap in Computing employment

In a news article (Evans, 2017), a skills gap was identified in Oklahoma, and it was stated that there was a reasonably high percentage of employers who found it difficult to hire employees with the required skills and educational background for the job. This is a challenge that has been consistent through the years, as can be seen in Trauth et al. (1993) where it was shown that there was an "expectation gap" between the skills required by Information Systems employers and the skills imparted to students in the Information Systems program. This gap led the authors of Trauth et al. (1993) to describe the Information Systems education system as lacking the ability to produce skillful and employable Information Systems (IS) professionals. They were able to prove their claim by carrying out qualitative and quantitative research involving IS professionals, professors, and recent graduates in New England in the US.

These skills gap claims are backed up by a report of the Commission on Higher Education and Employability (2018), published by the New England Board of Higher Education. On the issue of skills gap, higher education institutions and employers seem to hold differing opinions on the job readiness of college graduates. While 96% of higher institutions believe that their graduates are well-equipped for the job market, only 11% of employers believe that college graduates possess the skills they require for the workplace.

As mentioned earlier, this issue extends beyond the shores of the United States, as we can see in Llorens et al. (2013) where the authors were able to show that, in Spain, a gap was identified between the Information and Communication Technology (ICT) industry skill requirements and the skills taught by the ICT curricula. The data for this study was gathered using surveys. An article (Jacobs, 2015) discussing some research work carried out at a University of Technology in South Africa shows that there is a demand on universities to produce employable graduates. This is a reasonable demand, given that only about 37% of the graduates were employed without them undergoing any form of curriculum adjustment to enhance their employability.

These trends are visible in the computing fields, as well. Despite the continual rise of enrollment rates in computing departments as seen in the Computing Research Association (CRA)'s Taulbee surveys (Zweben & Bizot, 2016 – 2019), there still seems to be a shortage of suitable graduates with the required skills for industry (Exter et al., 2018; TechRepublic, 2016). A study was carried out by Exter & Turnage (2012) where data was collected through semi-structured interviews of experienced computing professionals. A statement made by one of those computing professionals reads thus: "Preparation in school was a nice thing but wasn't necessarily what I really had to know to do [the job]; there were a lot of things I still had to learn" (Exter & Turnage, 2012, p. 1). This further strengthens the claim that there is a gap between what is learnt in school and what is required at work.

2.2.2 Bridging the Skills gap in computing employment

There have been several research endeavors and projects that aim to improve the skill sets of computing students, especially within higher education. An example is seen in Exter et al. (2015) where a new multi-disciplinary program, which infused liberal arts and project-based learning into a computing program, was evaluated. The purpose of this approach to program design was to instill competencies in students as they learn through experience. According to Waguespack et al. (2019, p. 2), competency is defined as “a model of knowledge skillfully applied in task and disposed to an ethic of professionalism”. Several studies have also identified project-based learning as a technique that would potentially help students to develop needed skills and competencies (Chookittikul & Maher, 2011; Dragoumanos, 2017; Exter et al., 2018; McCrone et al., 2019). Other techniques such as the development of competency-based programs and collaborations between the industry and academia have also been employed in the computing field (Bannikova et al., 2018; Blackburn et al., 2016; Riel et al., 2016; Van Epps et al., 2016). In a similar vein, we see the rise of a company named Andela that builds distributed engineering teams through the recruitment of skilled software engineers from Africa. These teams work as engineers with companies ranging from start-ups to Fortune-500 companies in order to alleviate the shortage of skilled software developers in the computing industry (LinkedIn, 2020; Wikipedia, 2020). These endeavors support the observation that there is truly a skill shortage in the sphere of computing employment and that the shortage of skills leads to the unemployability of job seekers.

2.3 GENDER AND RACIAL GAPS IN COMPUTING EMPLOYMENT

2.3.1 The concept of underrepresented minorities in computing

2.3.1.1 Underrepresentation in computing employment

Over a decade ago, it was discovered that the ability of the United States (U.S.) to compete economically on a global scale is highly limited because of the US’s inability to develop her science and engineering workforce. The STEM [science, technology, engineering, and mathematics] workforce is generally acknowledged to have a substantial effect on America's capacity to compete globally (Gawlowicz, 2007; Varma, 2018). Hence, we see that the former U.S. President – Barack Obama – announced the "Computer Science for All" initiative in 2016 to build computing skills and computational thinking abilities in students, so that they can contribute to the digital economy (Smith, 2016). President Barack Obama (Obama, 2007) also declared that increasing the diversity in the STEM workforce would make the USA more competitive.

The major demographic groups that have been identified as being underrepresented in the computing field (computing studies and computing workforce) are: Women, Blacks, Hispanics, American Indians, and Alaska Natives (Google & Gallup, 2016; Varma, 2018; Funk & Parker, 2018). It is important to diversify the computing workforce because of many benefits that would accrue as a result. Women and the other racial minorities would contribute new innovation from their unique perspectives; their presence in the computing workforce would also be more reflective of the computing/technology user base than if they were excluded from the workforce (Varma, 2018). These benefits, among others, would culminate in economic growth for the computing field and for the USA as a whole (McClelland, 2001).

How has the computing workforce landscape looked like (in terms of its diversity) over the years?

Even though there have been numerous efforts to increase diversity in the STEM workforce, there hasn't been a very significant difference in the STEM workplace demographics in spite of the increase in representation of women and racial minorities in the workplace.

In 2018, women constituted 51% of the population and 46% of the civilian labor market in the United States (Varma, 2018). However, they only made up 29% of the STEM workforce and 24% of the computing workforce (Varma, 2018; Funk & Parker, 2018). According to the Pew Research

Center (Funk & Parker, 2018), we see (in Figure 2.1) that the representation of women in computing jobs has declined since 1990.

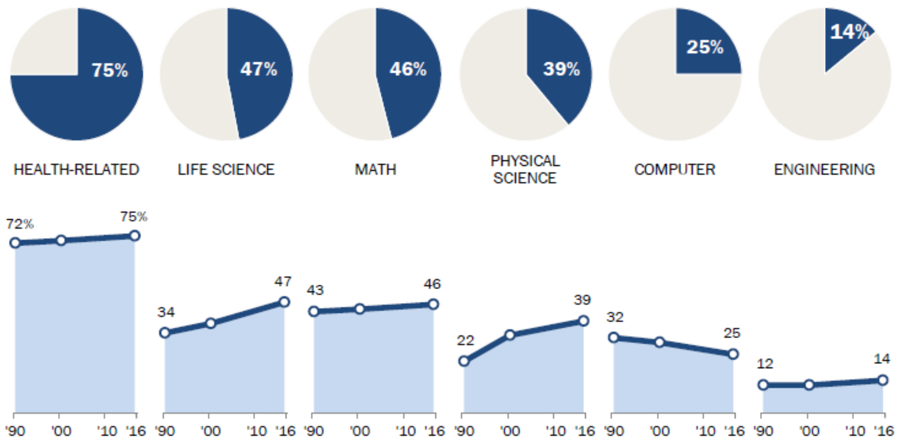

Figure 2.1 Representation of women in STEM occupations from 1990 to 2018

(Source: Funk & Parker, 2018)

The decline of the representation of women in computing jobs from 1990 to 2018 is particularly concerning because within that same time frame (1990 to 2014/2016), there has been more than a 338% increase in computing workers.

The representation of women in the computing workforce remained almost steady from 2018 up till 2019 as can be seen in Figure 2.2.

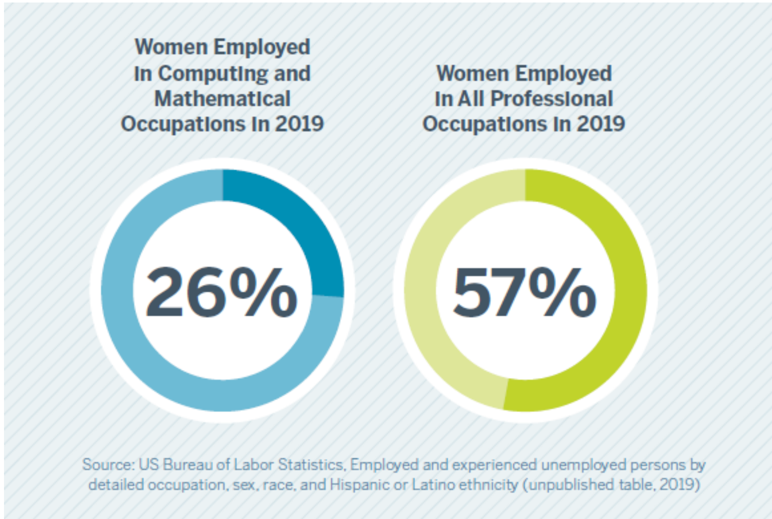

Figure 2.2 Representation of women in computing occupations in 2019

(Source: DuBow & Gonzalez, 2020)

Despite their low representation, women in the STEM workforce have a higher representation than that of the racial minorities. While Whites made up about 70% of STEM workers in 2013, Hispanics, Blacks, and American Indians or Alaska Natives recorded a much lower participation in the STEM workforce. Hispanics made up about 6% of the STEM workforce. Comparably, Blacks made up about 9% of the STEM workforce. The American Indians or Alaska Natives made about 0.2% of STEM workers. Though the racial representation of underrepresented racial minorities is still very low, these statistics show that there has been an improvement from what obtained in 1990 when STEM workers were 83% white, 4% Hispanic, and 7% black (Varma, 2018; Funk & Parker, 2018).

2.3.1.2 Underrepresentation in the pursuit of computing skills development

As much as the focus is on improving representation of the traditionally underrepresented minorities in the computing workforce, it is important to note that an underrepresentation of women and other racial minorities in computing studies would directly impact their underrepresentation in the workforce. We can see in Figure 2.3 that the top reason for the underrepresentation of Blacks and Hispanics in STEM jobs is the lack of access to quality STEM education.

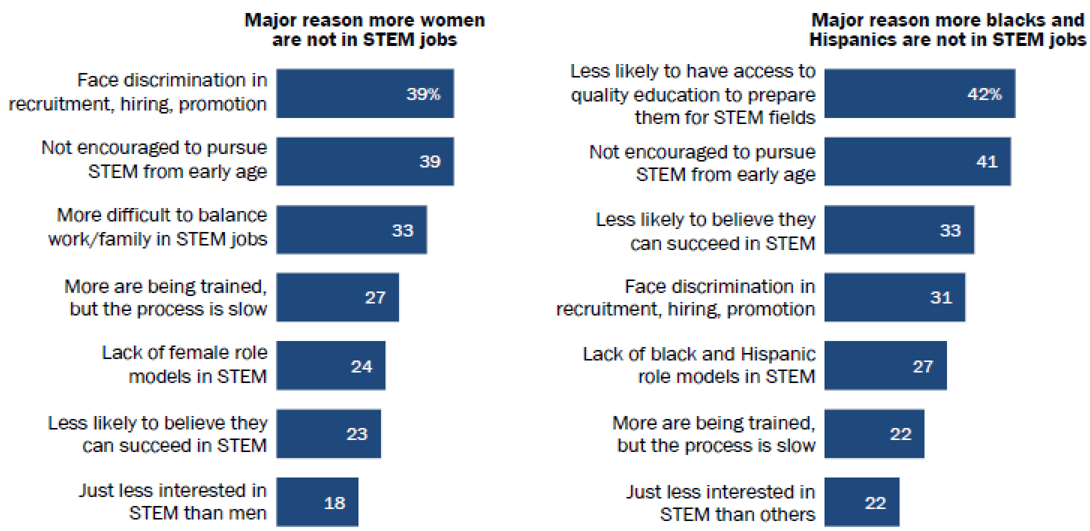

Figure 2.3 Reasons for underrepresentation of women and racial minorities in computing jobs (Source: Funk & Parker, 2018)

Unfortunately, the reverse does not necessarily hold. That is, a higher representation of underrepresented minorities in computing studies does not guarantee a commensurate rise in their representation in the computing workforce. We can see in Figure 2.4 that only 38% of women who had a college degree in computing went on to work in the computing field. However, increasing representation at the educational level would increase the possibility of higher representation in the workforce. Therefore, as the representation (or underrepresentation) of women and minorities is studied at the computing workforce level, it also needs to be studied at the computing studies level.

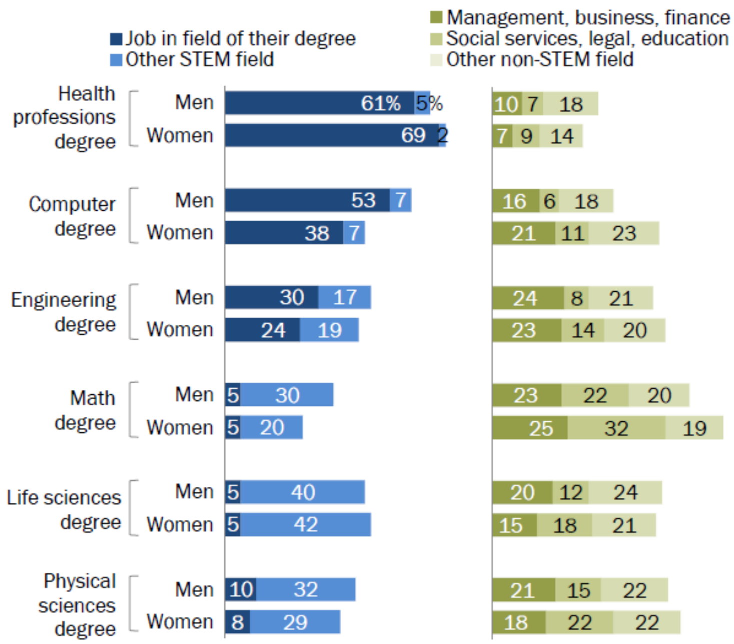

Figure 2.4 Relationship between STEM studies and STEM workforce retention

(Source: Funk & Parker, 2018)

How has the computing studies landscape looked (in terms of its diversity) over the years?

According to research by Google & Gallup (2016) about CS learning in the 7th to 12th grade, they discovered that "Male students (59%) are more likely than female students (50%) to say they have ever learned CS, and they are more likely to pursue opportunities to learn CS outside of the classroom", even though CS learning opportunities are equally available to both male and female students. This could be as a result of males having a higher level of confidence ("very confident": 65% male, 48% female) that they can learn CS, and a higher level of interest ("very interested": 34% male, 16% female) in learning CS (Google & Gallup, 2016). This result aligns with the underrepresentation of females in the computing field. If less females are interested in or confident about learning CS, it is no surprise that few women actually learn CS and proceed into the computing workforce.

In the U.S. post-secondary space, we see a similar trend: a smaller percentage of computing degrees are awarded to women (20% compared to 80% males) (Whitney & Taylor, 2018). This percentage rose to 21% in 2019 (DuBow & Gonzalez, 2020). Also, the representation of women as computing degree recipients in 2015 is not much different from their representation in 1987 while the representation of males more than doubled. Figure 2.5 shows women as a gender minority in computing college education from 1985 up until 2015.

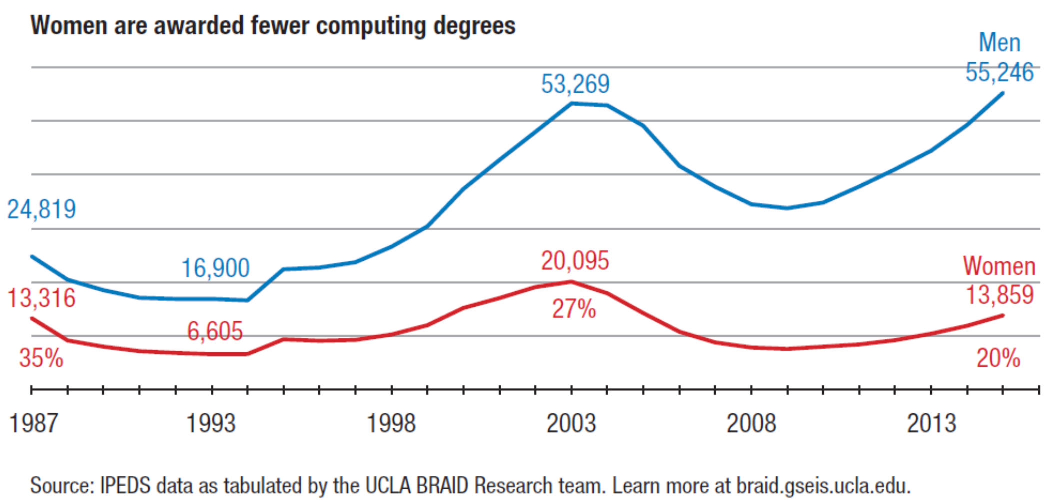

Figure 2.5 Representation of women in computing postsecondary education completion

Source: Whitney & Taylor, 2018

A contrary trend was found for the underrepresented racial groups. Black and Hispanic students, who are traditionally underrepresented in computing, were found to be more interested in learning CS (“very interested”: 33% Blacks and Hispanics, 20% White) than their White counterparts (Google & Gallup, 2016). Similarly, Black students exhibited more confidence in their ability to learn CS (“very confident”: 68% Blacks, 56% White, 51% Hispanics), than White or Hispanic students (Google & Gallup, 2016). Naturally, higher interest and confidence in learning CS should result in an increase in the representation of Blacks and Hispanics in computing studies, but this is not the case with the Black and Hispanic minority. Google & Gallup (2016, p.17) report that “Blacks and Hispanics continue to be underrepresented in CS fields. This indicates the factors that contribute to the underrepresentation of racial and ethnic minorities in CS fields go beyond student interest and confidence in learning CS”.

Between 1987 and 2015, growth has been recorded in the representation of racial minorities as computing degree recipients (Whitney & Taylor, 2018). The representation of Hispanics, Blacks, and Native Americans at least doubled between 1987 and 2015 but their representation is still very low compared to the 53% representation of Whites in 2015. This is pictured in Figure 2.6. The trends were still similar, even up till 2019. Among both men and women, there are more white computing degree holders than other races, with Blacks and Hispanics remaining in the minority (DuBow & Gonzalez, 2020). This can also be seen in Figure 2.7.

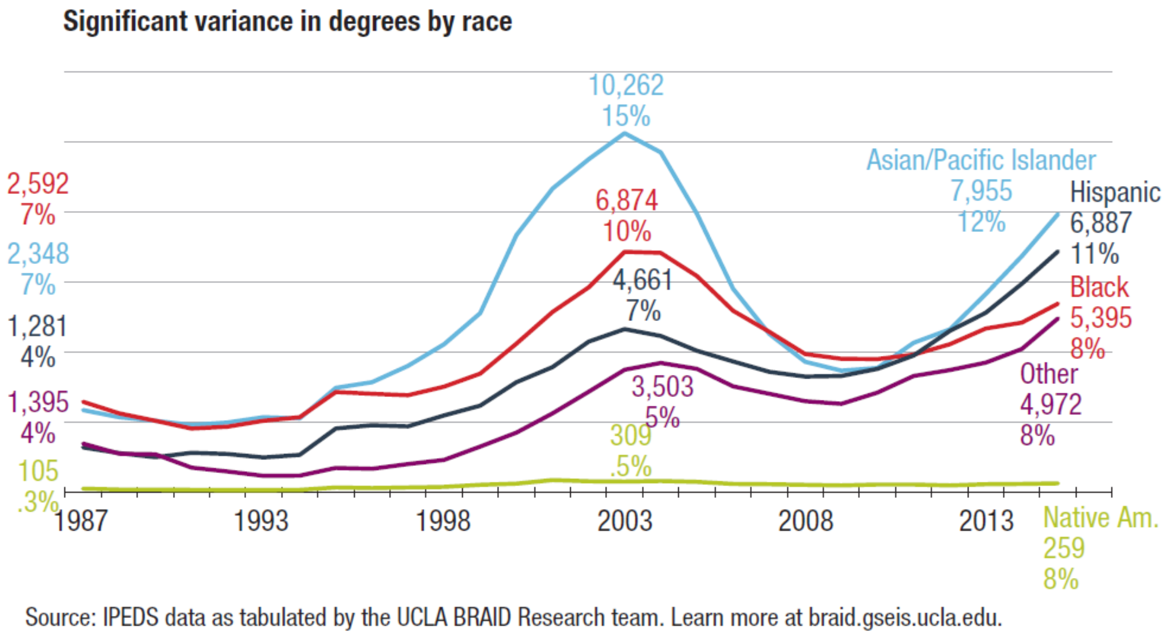

Figure 2.6 Representation of racial minorities in computing postsecondary education completion

(Source: Whitney & Taylor, 2018)

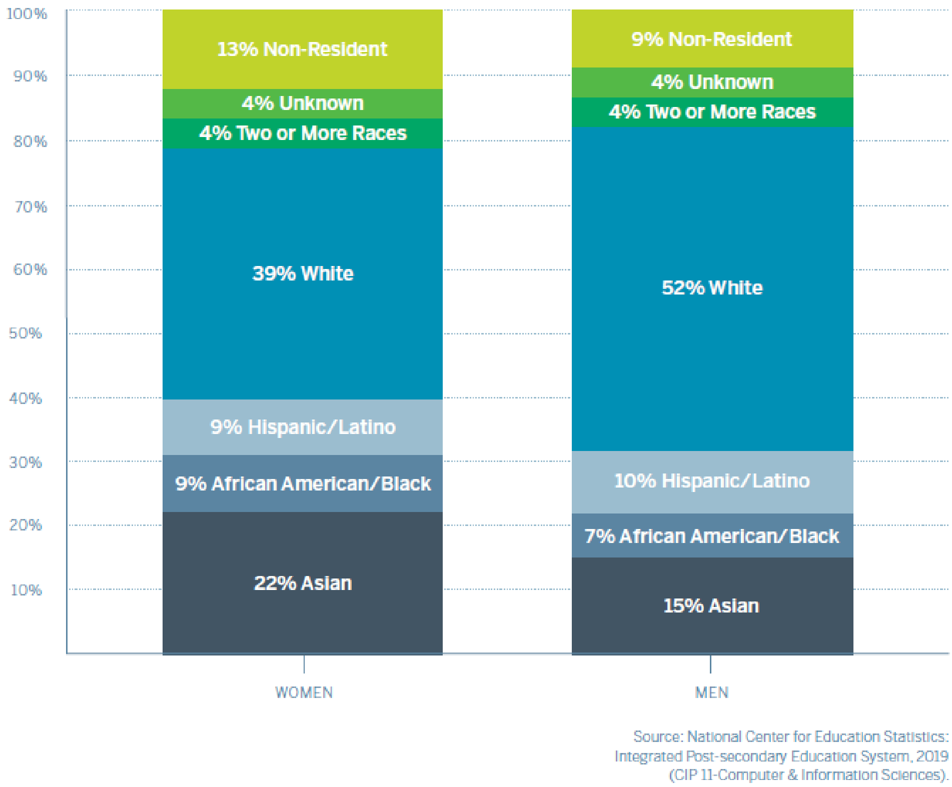

Figure 2.7 Computing degrees awarded in 2019 by race and gender

(Source: DuBow & Gonzalez, 2020)

2.3.2 Barriers to representation of women and other racial minorities

It has been established that females, Blacks, and Hispanics are underrepresented in computing education and workforce (Google & Gallup, 2016; Varma, 2018; Funk & Parker, 2018). Researchers have discovered several factors that contribute to the underrepresentation of these minority groups.

2.3.2.1 Lack of access and exposure

Many underrepresented minorities have been found to have less access and exposure to CS than the overrepresented groups (Simard, 2009). Google & Gallup (2016) showed in their research that Black and Hispanic students are less likely than White students to use a computer at home for most of the week. They also reported that, compared to their White counterparts, there is a lesser likelihood of Black students having CS classes at their school. In addition, it was observed that female students have a lesser likelihood of exposure to CS learning opportunities than their male counterparts. This shows that the exposure of underrepresented minorities to computing (both at home and in school) is limited. Majority (73%) of Black people in STEM believe that limited access to quality education is a major reason for the underrepresentation of blacks and Hispanics in STEM. 53% of Hispanics, 52% of Asians, and 50% of Whites also hold this view (Funk & Parker, 2018). This is pictured in Figure 2.8. This seems like a valid barrier to representation in computing because there is a higher level of educational attainment required for STEM jobs, under which computing falls (Funk & Parker, 2018).

2.3.2.2 Lack of motivation

Since motivation is a key factor to getting things done, the motivation levels of the underrepresented minorities in computing have also been studied. Aish et al. (2018) emphasized the importance of motivation through external influences to increasing the participation of underrepresented minorities in computing. Google & Gallup (2016) found that only about 11% of women and 16% of Hispanics claim to have seen people like them “doing CS” on TV shows. Seeing people like them doing CS inspires underrepresented minorities to desire to achieve similar goals, but the statistics show that women and Hispanics have a very low percentage of exposure to role models and positive external influences. Many Americans blame the STEM workforce's lack of diversity on a lack of early support for females, blacks, and Hispanics to pursue STEM careers (Funk & Parker, 2018). Simard (2019) also states that a lack of role models contributes to the limited participation of underrepresented minorities in computing.

2.3.2.3 Lack of confidence

The confidence factor is key in taking on a field which is highly technical. Unfortunately, many underrepresented minority groups have not built a lot of confidence in their ability to succeed in computing. Female students are less confident (48% vs. 65%) in their ability to learn CS than male students are (Google & Gallup, 2016). There is not a high level of confidence from external sources either. Google & Gallup (2016) reported that "Male students are more likely to be told by a parent or teacher that they would be good at CS (46% vs. 27% being told by a parent; 39% vs. 26% being told by a teacher)". As seen in the percentages, there is a lower level of confidence for female students from their parents and teachers.

2.3.2.4 Societal perception of computing

Research has shown that many people see computing as a "masculine field" (Peckham et al., 2007; Simard, 2009). This could be the reason why more males are encouraged to study computing, as we saw earlier. Google & Gallup (2016) also identified the perpetual social perception that CS is for certain groups of people: White or Asian males. In addition, CS is perceived as having little to no social relevance (Peckham et al., 2007) which women seem to care more about. These perceptions all work together to draw more Whites and males into CS; and discourage more females, blacks and Hispanics from the CS field.

2.3.2.5 Environmental conditions of the underrepresented

Many times, underrepresented minorities in the computing classroom, and even in the workforce, do not feel comfortable in the majority-dominated environment. Because they are minorities, they usually experience tokenism (Simard, 2009). Some have been labeled according to existing stereotypes of them not belonging to the computing field. Some have felt isolated (Simard, 2009). Others have been ignored, excluded, or overlooked by teachers in the class and the university environment as a whole (Simard, 2009), and others have experienced discrimination in the computing workplace (Peckham et al., 2007; Funk & Parker, 2018).

It is obvious that some of the above-listed barriers are direct consequences of other barriers to representation. For instance, there is a correlation between an underrepresented student's lack of exposure to CS and their lack of confidence in their ability to learn CS (Google & Gallup, 2016). It is also important to note that the above-listed barriers are only a few of the existing barriers to the representation of underrepresented minorities in computing. Figure 2.8 shows a broader view of the barriers.

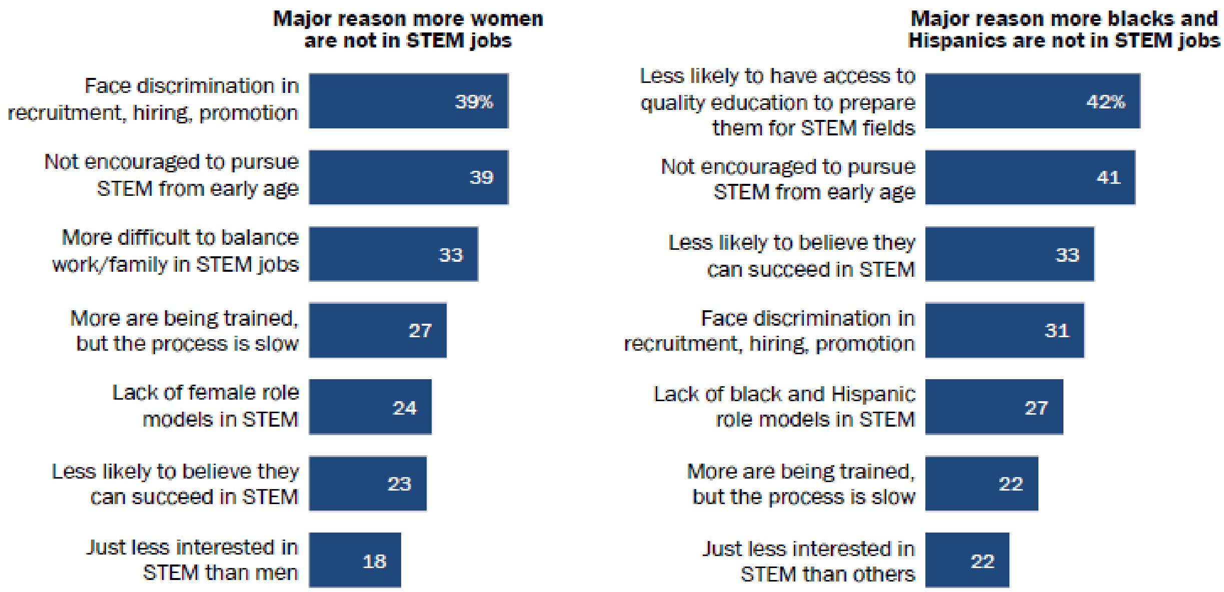

Figure 2.8 Reasons more women, Blacks, and Hispanics are not working in STEM

(Source: Funk & Parker, 2018)

Underrepresented minorities experience achievement/skill gaps, gender gaps, or race gaps in computing education and in the computing workforce because of the above reasons, among others.

2.4 BRIDGING THE GAPS IN THE COMPUTING FIELD

2.4.1 The need to bridge the employment (skills, gender, and racial) gaps in the computing field

Seeing that the computing field is inundated with male, White, and Asian workers (Peckham et al., 2007; Varma, 2018), a great strategy to fill up these vacant jobs is to attract traditionally underrepresented minorities into computing. According to research, attracting underrepresented minorities into computing studies would eventually result in the availability of a higher number of computing professionals for the workforce (McClelland, 2001; Peckham et al., 2007). If we had as many women, Blacks, and Hispanics in the computing workforce as we have men, Whites, and Asians, the demand for computing professionals would no longer be as high; hence providing a solution to the shortage of professionals in the computing workforce (Dasgupta & Stout, 2014).

It is also important that the underrepresented minorities are skilled for the available positions in the computing industry. Therefore, in recruiting minorities, attention needs to be paid to the computing studies programs/institutions that impart the relevant skills to them; recruit the minorities into such programs, and then produce adequately skilled underrepresented members of society to fill up the computing positions.

2.4.2 Efforts to bridge the employment gaps in the computing field

To achieve the aim of broadening participation in computing, many researchers have set out to study and provide strategies to overcome the barriers to representation of the underrepresented minorities in the computing field. Different solutions have been targeted towards different stages of life (or different stages along the computing pipeline) that the underrepresented minorities go through.

2.4.2.1 Early Stages (K-12, High School education)

To attract girls into the STEM field (of which Computing is a part), setting up partnerships between K-12 schools and science museums have proven to be helpful. From the discussion of barriers to representation, which was discussed earlier, we see that women show a lack of interest in computing whenever they do not see a social or altruistic relevance of computing studies. Exposing K-12 girls to science museum helps to draw girls into STEM by showing them the real-world relevance of STEM studies (Dasgupta & Stout, 2014). Creating opportunities for girls to get directly involved in solving real-world problem through informal STEM learning is a strategy that was recommended by Dasgupta & Stout (2014). A collaboration between K-12 schools and STEM departments in Higher education was also recommended by Dasgupta & Stout (2014). These would give the underrepresented K-12 students a chance to meet role models who look like them, thus increasing their interest and confidence to pursue a career in STEM. Margolis et al. (2012) also discuss their "Exploring Computer Science" program which is a K-12 / University collaboration aimed at broadening participation in CS at the high school level. The STARS (Students and Technology in Academia, Research, and Service) Alliance also employs the strategy of collaboration to broaden participation in computing (Dahlberg et al., 2011; Brown, 2016)

Efforts to broaden the participation of early stages underrepresented minorities in computing include: The Scalable Game Design project (Webb et al., 2012) targeted at middle school students to motivate their interest in computing and to develop their capacity for computational thinking. This project achieved a high level of participation of females and underrepresented minority students.

Some other products and strategies that have been developed to encourage underrepresented minorities in the early stages include EarSketch; a hybrid platform (included within a high school Computer Science Principles course) that combines computing with music (Freeman et al., 2015), the NSF-supported Mobile CS Principles (Mobile CSP) course; a high-school level introductory CS course that fosters student engagement, gets students to build mobile applications that are "socially useful", and targets underrepresented minorities; all to broaden participation in CS (Hoffman et al., 2019).

2.4.2.2 Intermediate stages (Enrollment and persistence in Higher education)

According to Dasgupta & Stout (2014), increasing participation at this stage revolves around creating a sense of social belonging for the underrepresented minorities. The authors suggested that, for women, exposure to fellow women in their field (both peers and mentors) would result in increased representation. Research by Alvarado & Judson (2014) alludes to the effectiveness of this strategy in broadening participation of women in Computer Science. They show us that attending Grace Hopper Conference (a conference for women in CS) before a student declares their major is highly influential on their choice of a major and career. This strategy was employed in Harvey-Mudd college, which was able to almost quadruple (from 12% to 40%) the percentage of women majoring in CS at the college (Alvarado et al., 2012).

Lamar University embarked on a program named INSPIRED (Increasing Student Participation In REsearch Development) to attract and retain women and other underrepresented minorities in CS. They employed the provision of peer support, mentors and role models, and the exposure of undergraduates to research and useful applications of CS (Doerschuk et al., 2009).

Participation in undergraduate research early in their undergraduate program helps to produce skilled professionals within the STEM field (Ordonez et al., 2020). On this basis, Ordonez et al. (2020) designed a summer research program targeted towards attracting Hispanic women in STEM into a computational research career.

Through curriculum modification, outreach, and the provision of community for underrepresented minorities at 15 universities, the BRAID (Building, Recruiting, And Inclusion for Diversity) scheme has been proven to make a lot of difference in increasing representation of the traditionally underrepresented students at the intermediate level (Whitney & Taylor, 2018).

Exposure of underrepresented minorities in the intermediate stages to role models has also helped to improve recruitment, retention, and overall broadening participation efforts as seen in Aish et al. (2018).

Organizations such as the National Center for Women & Information Technology (NCWIT), the Association of Computing Machinery (ACM-W), Anita Borg Institute (ABI), Computing Research Association (CRA-W), Center for Minorities and People with Disabilities in Information Technology (CMD-IT), among others, have been established to increase the representation of women and minorities in computing studies and beyond, and they have recorded success thus far (DuBow et al., 2016; Whitney & Taylor, 2018).

2.4.2.3 Across the Computing Pipeline

"Georgia Computes!" is an NSF-funded scheme that aims to broaden participation across the entire computing pipeline i.e., igniting student interest and improving the quality of education from the early stages, and increasing college enrollment, persistence, and further education in computing within the state of Georgia (Bruckman et al., 2009). As at the time Bruckman et al. (2009) published their paper, it was reported that there were significant positive results at the different stages of the pipeline, but the impact across the entire computing pipeline was not yet apparent.

CAITE (The Commonwealth Alliance for Information Technology Education) also established a wide range of programs targeted at the increasing participation of the traditionally underrepresented minorities across K-20 landscape (i.e., along the entire computing studies pipeline). CAITE aims to attract these students, prepare them for, and support them through a high school – community college – university – graduate school educational pathway. The inclusion of community colleges in CAITE's strategy is vital because community colleges provide access to the underrepresented communities (Adrion et al., 2008)

2.4.2.4 Alternative Routes

Apart from targeting the traditional education pipeline into the computing workforce to broaden participation of underrepresented minorities, Fealing et al. (2015), in their research, discussed the pathway model which states that there are other routes that lead into the STEM workforce. The pathways model goes further to state that there is a possibility that the solution to the diversity problem in STEM is not to graduate more people out of the traditional pipeline but to place more value on the alternative routes that are usually taken by underrepresented minorities (Fealing et al., 2015).

Blaney (2020) also emphasized the need to shift our broadening participation focus from the traditional pathways into computing college programs to other pathways such as the upward transfer students (students who transfer from community colleges to 4-year computing college programs), since there is a high percentage of underrepresented students among the upward transfers.

These different broadening participation efforts at different stages have recorded success in their different contexts of application. However, beyond the immediate contexts, it is important to determine the impact of these efforts beyond getting to the end of the computing education pipeline (i.e., beyond earning a computing degree). It is vital to observe the impact of these efforts on the eventual employment outcomes of the underrepresented minorities.

2.4.3 Defining Success: Measuring the impact of these efforts on computing employment

Looking ahead to the ultimate goal, which is successful computing employment, how successful are these efforts/schemes to broaden participation? Apart from the immediate results on academic performance, recruitment, persistence across the computing pipeline, self-efficacy, etc., what is the impact of these schemes on the eventual employment outcome of the underrepresented minorities?

Also, since these different solutions have different contexts of application (e.g., some solutions are geared towards early stages, some towards intermediate stages, some towards providing alternate pathways to employment for the underrepresented community, and others across the entire computing pipeline), what is the basis/foundation for these solutions that are targeted at these different contexts? Is there a reason why a specific solution (e.g., INSPIRED) was targeted towards the undergraduate education landscape? Or is the "Broadening Participation in Computing" (BPC) community coming up with solutions that might work to address some of the barriers to representation without basing the solution on hard evidence that shows that their effort will result in successful employment outcomes for the underrepresented minorities in their context? Could this be why the percentage of women in the STEM workforce has declined since 1990, and the percentage of racial minorities in the computing workforce has hardly broken into the double digits?

How about a more targeted curation of solutions? What about a situation where researchers can use available data regarding the most effective aspects of education (and the aspects of education that require the most intervention) for achieving a successful employment outcome for the underrepresented minorities in computing? If such information is available, wouldn't there be a high possibility that the BPC schemes would be developed based on an existing need (as reported by the data) for those specific schemes, and that they would result in the desired employment outcomes for the underrepresented minorities?

Instead of trying to fix what does not work, what about studying the existing success stories (and pathways) of underrepresented minorities, identifying what has worked for them, and using the discovered information as building blocks for future solutions to replicate the successes that have been studied? If a person is successful in achieving a great computing employment outcome, it would mean that they were able to surmount the skills gap issue. It would mean that their employer saw them as fit for the job: skill-wise, and otherwise. So, for the underrepresented minorities in computing who have achieved successful computing employment outcomes, what are the educational choices they made that enabled them to surmount the skills gap, gender gap, and race gap in order to attain successful computing employment?

Answering this question would enable researchers and decision makers to have a clearer picture of what educational choices have worked historically for underrepresented minorities in computing, and consequently, researchers and decision makers would be able to more accurately target their BPC efforts to achieve optimal results.

2.5 WORKING BACKWARDS: PROVIDING A MORE ACCURATE PICTURE OF THE COMPUTING EDUCATIONAL AND EMPLOYMENT LANDSCAPE

2.5.1 An alternative solution to broadening participation in the computing workforce

Many BPC efforts in the existing literature have designed and deployed solutions aimed at removing one or more barriers to representation, after which the impact (of the solution) on the representation of underrepresented minorities is evaluated. This is what this research describes as "working forward". Using this strategy, several hypotheses are put forward and tested with the expectation that the employment gaps will be bridged.

Contrary to the "working forward" strategy, this research sets out to describe and test the "working backwards" strategy where the analysis begins with underrepresented individuals who have achieved successful employment outcomes. From this end point, this research works backwards to observe the pathways (educational decisions) that have resulted in those successes.

Because the "working backwards" strategy begins from a position of success, it is presented in this research as a strategy that promises to deliver a higher rate of success if the insight drawn from this strategy is used as a basis for future broadening participation efforts.

2.5.2 Why study pathways?

In order to paint a picture of what practically works (and does not work) for the representation of underrepresented in the computing workforce, this research studies the educational pathways of successful individuals who identify as underrepresented minorities in computing. The educational pathways refer to the educational choices made by individuals to acquire the skills they require to successfully gain computing employment. According to the NSF Award #1104195 document on the topic "Pathways to Cybersecurity and Information Assurance Careers" (National Science Foundation, 2011), the CCTI (College and Career Transitions Initiative) defines a career pathway as "an articulated sequence of rigorous academic and career courses, beginning in the ninth grade and leading to an associate degree, and/or an industry-recognized certificate or licensure, and/or a baccalaureate degree and beyond". We also find a definition of "college and career pathways" in

MDRC (2015). The “pathways” are defined in the report as “a range of models or approaches that attempt to create a clear path for students to follow to attain an educational and occupational goal, while learning the skills they need to succeed in both domains”.

Identifying the most relevant pathways (leading to successful employment outcomes for the underrepresented minorities) would inform future efforts to increase representation by investing in the educational pathways that promise to yield the highest returns in broadening participation.

The educational pathways (of the successful underrepresented minorities) are studied because of the vital role that they play in the eventual employment outcomes of individuals in the workforce. Negoita & Dunham (2013) show that many young individuals are unable to successfully enter the workforce because they are either unaware of or unable to access certain educational programs that have the potential to prepare them for a successful career.

Thus, identifying the more impactful computing educational pathways will constitute vital information for young people who identify as underrepresented minorities in computing.

2.5.3 Traditional Pathways

Bachelor’s degrees (4-year colleges) (Bowman, 2018; NCES, n.d.) are one of the major sources of computing professionals for the computing workforce. About 40% of people that were employed in professional jobs (under which computing falls) in 2017 and 2018 had college degrees (NCES, n.d.).

According to Kannankutty (2007), there are multiple unique pathways that people can take to enter their careers. The author also identified distinct stages through which most people journey. These set of stages was referred to as a "continuum" that people generally progress through. The continuum describes the traditional pathway into employment. Figure 2.9 is a diagrammatic depiction of this continuum. As seen in Figure 2.9, people would typically go through pre-college education (high school, community college, etc.) and then transition to an undergraduate education at a college/university. After earning their undergraduate degrees, they would typically enroll in graduate school (for a Master's or Doctoral degree), and then transition into the workforce. However, in reality, this is not how many people progress into the workforce.

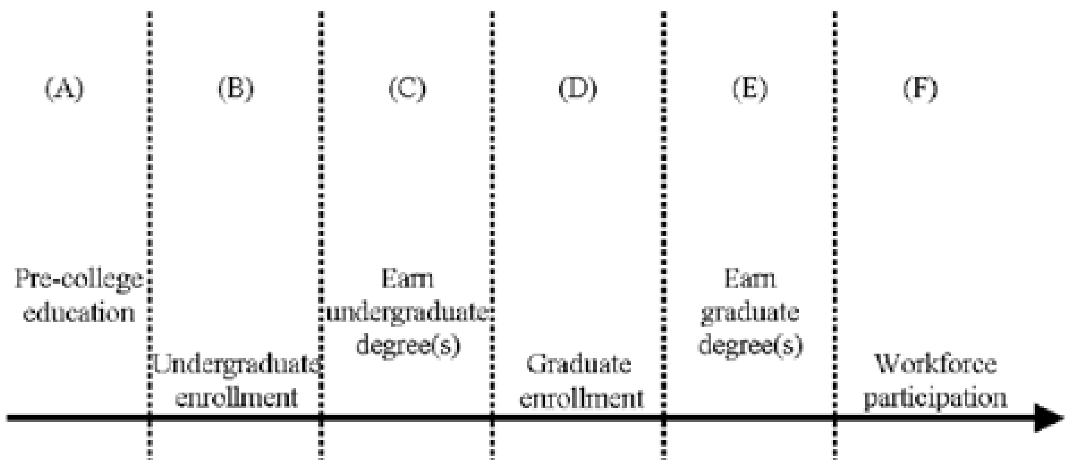

Figure 2.9 The Education and Workforce continuum

(Source: Kannankutty, 2007)

2.5.4 Alternative Pathways

The other 60% of people employed in professional jobs must have entered the workforce through alternative pathways. As seen in literature, some alternative pathways into the computing workforce include certifications (Olagunju & Zongo, 2010; Randall & Zirkle, 2005), coding bootcamps (Price & Dunagan, 2019; Seibel & Veilleux, 2019; Joshi, 2019; Lee et al., 2019), undecided students, community colleges, and "end user to end user programmers" (Lehman et al., 2018).

Carreon (2019) published a proposed pathway for high school students to enter into the computing workforce. The pilot program is to be a 4-year pathway that prepares high school students for entry-level software development jobs without a bachelor's degree. This 4-year pathway includes two years of relevant coursework in high school, one year of community college, and then one year to bag an associate's degree in computer science.

As seen in the state of Mississippi, coding bootcamps provide an effective alternative pathway to computing employment. As a response to the requirements of employers in Mississippi, and in an attempt to narrow the skills gap in the state, a non-profit organization called Innovate Mississippi (InnovateMS, 2020) set out to widen the pool of skilled computing professionals in the state by establishing the Mississippi Coding Academies (MCA). Many learners from underrepresented groups attend under-funded K-12 school districts (PCR&EF, 2018), and are often unprepared for college. They may lack a formed identity that provides them the confidence to pursue higher education (Verdin et.al., 2018; Krupnick, 2018). The MCA model provides an alternative pathway to post-secondary skills development and is a proven pathway for learners who have not formed identity with higher education.

The Bureau of Labor Statistics (2020) predicts a 12% growth in computing occupations from 2018 to 2028, "much faster than the average for all occupations." With the demand for computing talent increasing faster than the supply of potential hires, the Mississippi Coding Academies works with industry partners to identify pathways to hiring technical employees "skilled through alternative routes (STARs)" (Blair et.al., 2020).

2.5.5 Traditional and Alternative Pathways

Occasionally there exists a hybrid pathway, consisting of both traditional and alternative educational endeavors. An example was seen in Waters (2008) where the technology coordinator for Irvington High School in California discussed the idea of "seamless pathways" between the students at his school and the computing industry.

A partnership between Irvington High School and Ohlone College (a community college) was created to develop such a pathway program where students can get on career tracks into computer network technologies, multimedia and graphic design, or software engineering. This pathway involves structured training in high school in the chosen track, then enrollment in a community college, and finally, a transfer to a university. A partnership is also mentioned in Waters (2008) between Irvington, Ohlone, and San Jose State University where the pathway looks like this: 2 years of Cisco Networking Academy training in high school + 2 years in community college + 2 years at university.

2.5.6 Why study pathways for Women and Black people in Southern USA?

Historically, Blacks and women have been economically disadvantaged (Weller, 2019). Blacks continue to experience higher unemployment rates, as reported by the U.S. Bureau of Labor Statistics in 2019 (Weller, 2019). They also experience unfavorable labor market conditions such as low pay, job instability, long wait before getting a job, among others (Weller, 2019). Similarly, women face disadvantages in the job market when compared to men. White women earn less than their male counterparts. Since black people earn less than their white counterparts, black women earn even less than white women (Weller, 2019).

The southern region of the USA has also continued to lag behind the rest of the nation, economically. The states that make up the Southern region include: Mississippi, Louisiana, Alabama, Georgia, South Carolina, North Carolina, Virginia, West Virginia, Oklahoma, Arkansas, Tennessee, and Kentucky (Nunn, 2019). Since the South had been declared as “America’s Economic Problem Number 1” in 1938, the South has struggled to regain its economic prosperity (Kromm, 2011; Nunn, 2019).

Women are the economic drivers of families in Mississippi (MS) as cited by *Women Driving Change: A Pathway to a Better Mississippi.* Black women live at the intersection of a legacy of race and gender bias; 36.2% live in poverty. 80% of black mothers in MS are the primary source of economic support for the family, yet 41% are in low-wage jobs. They face unique challenges to education and training. "Nearly four in 10 MS women working in low-wage jobs are supporting children under the age of 18; of those women in low-wage jobs with children, over two-thirds are responsible for supporting children on their own." Those making the median of $26K must spend nearly 40% on childcare.

It is apparent that blacks and women in the South are at a double disadvantage in relation to employment outcomes; studying what pathways have resulted in successful employment outcomes will provide a body of knowledge useful for making informed decisions regarding the improvement of the employment outcomes for blacks and women in Southern USA.

CHAPTER III

METHODOLOGY

3.1 AIM AND RESEARCH QUESTIONS

The aim of this research is to study the educational pathways of traditionally underrepresented persons (blacks and women) who have successfully transitioned into the computing workforce and identify the factors that have contributed to their retention, persistence, and success in attaining computing employment. Similarly, the aim is divided into the following research questions:

1. What are the common themes across educational pathway experiences that emerge from the analysis of computing professionals' data across racial and gender dimensions?
2. Which of these common experiences result in successful long-term (greater than 3 years) employment outcomes in the technology sector for women and for blacks?
3. How do the findings of this study inform national investment in broadening participation efforts that seek to increase racial and gender diversity in the computing workforce?

3.2 DESCRIPTION OF DATA

In order to achieve this aim and answer the above questions, a longitudinal study of computing professionals who studied or worked (or are currently working) in Mississippi (and surrounding Southern states) is carried out. Their identification as computing professionals is based on the definition of a computing job as stated in sections 2.1.1 and 2.1.2. The educational and work history of these computing professionals is extracted from LinkedIn. "LinkedIn is an American business and employment-oriented online service that operates via websites and mobile apps. Launched on May 5, 2003, the platform is mainly used for professional networking, and allows job seekers to post their CVs and employers to post jobs" (Wikipedia, 2021). This data extraction is IRB-approved (See Appendix A: IRB Approval letter).

Given this data, the pathways of the data subjects are studied especially along the race and gender lines, in order to answer the questions listed above.

The longitudinal data from LinkedIn consists of 303 rows of data and 77 columns. This dataset contains raw information directly from LinkedIn, and thus cannot be shared for privacy purposes.

3.2.1 Categorization of LinkedIn Data

The educational and employment history of each individual is then categorized. This dataset is also stripped of identifiers.

3.2.1.1 Categorization of Educational History

- For every individual in the dataset:
    - Each educational experience is categorized as either a **computing** degree or **non-computing** degree.
    - Each educational institution is also categorized as either a **R1** (Doctoral Universities-Very High Research Activity), a **R2** (Doctoral Universities-High Research Activity), a **D/PU** (Doctoral/Professional Universities), an **M1** (Master's

Colleges and Universities – Larger programs), an **M2** (Master's Colleges and Universities – Medium programs), an **M3** (Master's Colleges and Universities – Smaller programs), a **BC** (Baccalaureate colleges), a **BAC** (Baccalaureate/Associate's colleges), an **AC** (Associate's colleges), an **SFI** (Special focus institutions (2yr, 4yr)), or a **TC** (Tribal colleges) institution, using the Carnegie classification. That constitutes 11 attributes of the **Carnegie Institution ranking** variable.

  - The Carnegie classifications are carried out using the lists found on Carnegie Basic Classification Description and Carnegie Classification of Institutions of Higher Education

3.2.1.2 Categorization of Employment History

- For every individual in the dataset:
  - For each employment experience, the **Employment Status** variable is created with possible attributes: **Employed**, **Unemployed**, or **Self-employed**.
  - Each employment experience is categorized as either Computing or non-computing, under the **Employment Type** variable. This categorization is done, based on what sections 2.1.1 and 2.1.2 and the existing literature (U.S. Bureau of Labor Statistics, 2021; Montandon et al., 2021; Indeed Editorial Team, 2021; US News, 2021; ComputerScience Staff, 2021; Panko, 2008; Learn How To Become, n.d.) characterizes as a computing job.
  - For each employment experience, the employer is categorized as either a **Fortune 500** or **Non-Fortune500** company, under the **Employer Rank** variable. The 2020

list of Fortune 500 companies (Someka Excel Solutions, 2020) is used to determine which companies were on the Fortune 500 list. Companies that are not on the list were categorized as Non-Fortune 500 companies.

- For each employment experience, the annual salary for job position is gotten from 3 established job websites: Indeed (Indeed, 2021), ZipRecruiter (ZipRecruiter, 2021), and Glassdoor (Glassdoor, 2021), and the average annual salary is calculated. This average salary is then categorized as either **Low** (less than $50,000), **Medium** ($50,000 to $100,000), or **High** (greater than $100,000), under the **Salary Level** variable. The list of job positions and their salary levels used for this research can be found at this link: Computing Salaries.xlsx

The resulting dataset consists of 301 rows and 120 columns and can be found at this link: Final Categorized Dataset.

3.2.2 Description of Final Dataset for Analysis

In order to prepare the data for analysis and eliminate empty data cells, the above dataset is further processed and reduced to 301 rows and 13 columns.

The variables in the dataset are described as follows:

- Race: This variable describes the race of the data respondent and it holds five possible attributes – **Black**, **White**, **Asian**, **Hispanic**, **Latino**, **Latina**.
- Gender: This variable describes the gender of the data respondent and it holds two possible attributes – **Male** and **Female**.

- Highest Degree Attained: This variable describes the highest level of education of the data respondent by grouping it into one of three categories, based on the individual's educational history. **HSL** (High School or Less) refers to an educational history that consists of only a high school degree or no education. **SC** (Some College) refers to an educational history that includes 1-3 years of college or a 2-year college program such as an Associates' degree. Finally, **BDH** (Bachelor's Degree or Higher) refers to an educational history that includes a Bachelor's degree, Master's degree, or a Doctoral degree. These categories can be seen in work done by Stoops (2004) and Castro & Coen-Pirani (2016).
- Traditional Degree: This variable holds the information about whether the data respondent holds any traditional degree ranging from an Associates' degree to a Doctoral degree. This variable holds 2 possible attributes: **Yes** and **No**.
- Alternative Degree: This variable holds the information about whether the data respondent holds any alternative degree including certifications, coding bootcamp degrees, or any other non-traditional degree. This variable holds 2 possible attributes: **Yes** and **No**.
- Computing Degree: This variable holds the information about whether the data respondent holds either a traditional or an alternative degree (or both) in a computing field or program. This variable holds 2 possible attributes: **Yes** and **No**.
- Highest Institution Ranking: This variable holds the rank of the highest-ranking institution attended by respondent. The possible attributes are the ranks which are listed in Section 3.2.1.1 above.
- Internship: This variable holds the information about whether the data respondent did an internship in any computing field. This variable holds 2 possible attributes: **Yes** and **No**.

- Current Employment status: This variable holds the current employment status for the data respondent (at the time of the data collection). The possible attributes are the employment statuses which are listed in Section 3.2.1.2 above.
- Time Elapsed before computing job: This variable holds information about the time that passed between the completion of the data respondent's last (most recent) computing education and the start of the first computing job. The last (most recent) computing education refers to either a traditional or alternative degree that was completed before and closest to the first computing job. The three possible attributes are **NW** (No Wait), **MW** (Moderate Wait), and **LW** (Long Wait). NW describes a situation where the data respondent attained computing employment before completion of most recent computing education; MW describes a situation where the data respondent attained computing employment within 1 year or less of completion of most recent computing education; and LW describes a situation where the data respondent attained computing employment over 1 year after completion of most recent computing education.
- Persistence in Computing field: This variable holds information about whether the data respondent spent at least three years in computing employment. This variable holds 2 possible attributes: **Yes** and **No**.
- Highest Computing Employer Ranking: This variable holds the rank of the highest-ranking employer (in a computing job) of the data respondent. The possible attributes are the employer ranks which are listed in Section 3.2.1.2 above.
- Highest Computing Salary Level: This variable holds the highest salary level in the computing employment history of the data respondent. The possible attributes are the salary levels which are listed in Section 3.2.1.2 above.

This final dataset with 301 rows and 13 columns can be found at this link: Final Dataset for Data Analysis

3.2.3 Limitations of LinkedIn Data

Using social media data for this research has its unique limitations.

LinkedIn, as a corporate social media platform, holds only the data that users supply. This means that the complete picture of a person's educational and employment experience might not be present on this social media platform. Therefore, the data used in this research might not be reflective of the whole picture of a person's education and career path.

Second, extracting the employment and educational data from LinkedIn is a slow and tedious process. This is because the educational and employment data on LinkedIn is in unstructured (text) form, and this unstructured data, from different LinkedIn pages, is extracted into a structured (tabular) form.

Third, the tediousness and time requirement of this data gathering process has produced a small dataset because of the time constraints and human resource constraint of this research. Because of the small size of this dataset, it is likely not representative of the general population. Similarly, the analysis of this small dataset will likely produce results that may not be generalizable to a larger population.

Even though there are significant limitations to using social media data, social media data provides a new opportunity to understand the pathways of women and blacks and to identify factors that result in their success in achieving and persisting in computing employment.

3.3 METHODOLOGY

3.3.1 Data Analysis Tools

In order to study the pathways of successfully employed women and black people in the computing field, some data analysis tools and techniques will be employed on the final dataset. Stata and the Orange tool for data mining are the data analysis tools to be used to extract intelligence from the dataset.

Stata is a general-purpose statistical software package developed by StataCorp for data manipulation, visualization, statistics, and automated reporting (Wikipedia, 2021). Orange is a component-based visual programming software package for data visualization, machine learning, data mining, and data analysis (Wikipedia, 2021)

3.3.2 Data Pre-processing for Stata

The 13 variables listed in section 3.2.2 are dichotomized and coded into dummy variables in Stata.

3.3.2.1 Pre-processing the independent variables

The independent variables in this dataset are Race, Gender, Highest Degree Attained, Traditional Degree, Alternative Degree, Computing Degree, Highest Institution Ranking, and Internship. 0 is assigned to the less desirable attribute and 1 is assigned to the more desirable attribute. Pre-processing is carried out on the independent variables as described below:

- Race: Whites, Asians, Hispanics, Latinos, and Latinas are merged into one category: Non-blacks. Blacks are one of the target groups of this work, therefore 1 is assigned to it. On the other hand, Non-blacks are the reference group.
    - Non-blacks: 0, Blacks: 1

- Gender: Females are one of the target groups of this work, therefore 1 is assigned to it. On the other hand, Males are the reference group.
    - Male: 0, Female: 1

- Highest Degree Attained: Each attribute of this variable is separated into its own variable and coded into dummy variables.
    - Bachelors' Degree and Higher: This variable holds the information about whether the data respondent possesses a bachelor's degree or higher. This variable holds 2 possible attributes: **Yes** and **No**. The "Yes" attribute of this variable is coded as 1 and the "No" attribute is coded as 0.
        - No: 0, Yes: 1
    - Some College: This variable holds the information about whether the data respondent possesses some college degree. This variable holds 2 possible attributes: **Yes** and **No**. The "Yes" attribute of this variable is coded as 1 and the "No" attribute is coded as 0.
        - No: 0, Yes: 1
    - High School: This variable holds the information about whether the data respondent possesses a high school degree. This variable holds 2 possible attributes: **Yes** and

No. The “Yes” attribute of this variable is coded as 1 and the “No” attribute is coded as 0.

        - No: 0, Yes: 1

    - The High school variable is set as the reference group, thus it is not visible within the data analysis model.

- Traditional degree: The “Yes” attribute of this variable is coded as 1 and the “No” attribute is coded as 0.
    - No: 0, Yes: 1
- Alternative degree: The “Yes” attribute of this variable is coded as 1 and the “No” attribute is coded as 0.
    - No: 0, Yes: 1
- Computing degree: The “Yes” attribute of this variable is coded as 1 and the “No” attribute is coded as 0.
    - No: 0, Yes: 1
- Highest Institution Ranking: The 11 attributes of this variable are compressed into 3 attributes:
    - Rank1: R1 (Doctoral Universities-Very High Research Activity), R2 (Doctoral Universities-High Research Activity), D/PU (Doctoral/Professional Universities)
    - Rank2: M1 (Master's Colleges and Universities – Larger programs), M2 (Master's Colleges and Universities – Medium programs), M3 (Master's Colleges and Universities – Smaller programs), BC (Baccalaureate colleges), BAC (Baccalaureate/Associate’s colleges)

- Rank3: AC (Associate's colleges), SFI (Special focus institutions (2yr, 4yr)), TC (Tribal colleges)
- Each new attribute (Rank1, Rank2, and Rank3) of this variable is separated into its own variable and coded into dummy variables.
  - Rank1: This variable holds the information about whether Rank1 is the rank of the highest-ranking institution attended by respondent. This variable holds 2 possible attributes: **Yes** and **No**. The "Yes" attribute of this variable is coded as 1 and the "No" attribute is coded as 0.
    - No: 0, Yes: 1
  - Rank2: This variable holds the information about whether Rank2 is the rank of the highest-ranking institution attended by respondent. This variable holds 2 possible attributes: **Yes** and **No**. The "Yes" attribute of this variable is coded as 1 and the "No" attribute is coded as 0.
    - No: 0, Yes: 1
  - Rank3: This variable holds the information about whether Rank3 is the rank of the highest-ranking institution attended by respondent. This variable holds 2 possible attributes: **Yes** and **No**. The "Yes" attribute of this variable is coded as 1 and the "No" attribute is coded as 0.
    - No: 0, Yes: 1
  - The Rank2 variable is set as the reference group, thus it is not visible within the data analysis model.

- Internship: The "Yes" attribute of this variable is coded as 1 and the "No" attribute is coded as 0.

- No: 0, Yes: 1

3.3.2.2 Pre-processing the dependent variable

The dependent variable was arrived at by pre-processing the last 5 variables in the final dataset: Current Employment status, Time Elapsed before computing job, Persistence in computing field, Highest Computing Employer ranking, and Highest computing salary level. These variables are pre-processed in a similar manner as the independent variables. 0 is assigned to the less desirable attribute and 1 is assigned to the more desirable attribute. Pre-processing is carried out on the last 5 variables as described below:

- Current Employment status: The "Employed" and "Self-Employed" attributes of this variable are merged into a single attribute – Employed. Thus, the three attributes of this variable are compressed into two and then coded into dummy variables.
  - Unemployed: 0, Employed: 1
- Time Elapsed before computing job: The "No Wait" and "Moderate Wait" attributes of this variable are merged into a single variable – Moderate wait. This is because "No wait" is also within the Moderate wait (<= 1year) time bracket. Thus, the three attributes of this variable are compressed into two and then coded into dummy variables.
  - Long Wait: 0, Moderate Wait: 1
- Persistence in computing field: The "Yes" attribute of this variable is coded as 1 and the "No" attribute is coded as 0.
  - No: 0, Yes: 1
- Highest Employer Ranking: The "Fortune 500" attribute of this variable is coded as 1 and the "Non-Fortune 500" attribute is coded as 0.

    - Non-Fortune 500: 0, Fortune 500: 1

- Highest Salary Level: The “Medium” and “High” attributes of this variable are merged into a single attribute – Good Income. Thus, the three attributes of this variable are compressed into two and then coded into dummy variables.
    - Low Income: 0, Good Income: 1

After the last 5 attributes of the final dataset are pre-processed, they are combined into an index. This index is referred to as the Success Index and its value is the sum of the dummy values that were assigned to the last 5 variables from the final dataset. These 5 variables are combined into a single index because they are the 5 indicators of successful computing employment outcomes as indicated in Section 2.1.3. The Success Index holds a value between 0 and 5, where 0 represents a very unsuccessful computing employment outcome and 5 represents a very successful computing employment outcome.

- In order to be able to carry out a logistic regression on the pre-processed data, the Success Index is converted to a new categorical variable (Employment Outcome) with two attributes: Unsuccessful and Successful. Success Index values of 1 – 3 are represented by the “Unsuccessful” attribute of the “Employment Outcome” variable, while the Success Index values of 4 and 5 are assigned to the “Successful” attribute of the “Employment Outcome” variable. The “Successful” attribute of this variable is coded as 1 and the “Unsuccessful” attribute is coded as 0.

The “Employment Outcome” variable is the dependent variable for the data analysis.

3.3.2.3 Resulting Features of Pre-processed Data

During the process of recoding all variables into dummy variables, the missing values are excluded from the final dataset. A common data sample is generated, leading to a reduction in size (compared to the size of the original dataset). The final pre-processed data is made up of 194 records (rows) and 13 variables (columns). Below is the descriptive statistics table for the pre-processed data:

Table 3.1 Descriptive Statistics of Data

Variable	Number of Observations	Mean	Standard Deviation	Min Value	Max Value
Independent Variables					
Race	194	0.180	0.386	0	1
Gender	194	0.180	0.377	0	1
Bachelors' Degree and Higher	194	0.938	0.242	0	1
Some College	194	0.052	0.222	0	1
Traditional degree	194	0.990	0.101	0	1
Alternative degree	194	0.392	0.489	0	1
Computing degree	194	0.995	0.072	0	1
Rank1	194	0.758	0.430	0	1
Rank3	194	0.041	0.199	0	1
Internship	194	0.242	0.430	0	1
Dependent Variable					
Employment Outcome	194	0.722	0.449	0	1

This table shows that across racial lines, about 18% of the data respondents are black while 82% are white; across gender lines, about 18% of the data respondents are female while the remaining 82% are male. Similarly, about 93.8% of respondents possess a bachelor's degree and higher, 5.2% possess some college degree, while the remaining 1% possess only a high school degree. 99% of the data respondents possess a traditional degree; 39.2% of the respondents possess an alternative degree; and 99.5% of all respondents possess a computing degree. Of all the data respondents, 75.8% attended a Rank1 institution, 20.1% attended a Rank2 institution, and 4.1% attended a Rank3 institution. Only 24.2% of all data respondents did an internship while the rest did not.

3.3.3 Data Pre-processing for Orange Data Mining

The final dataset from section 3.2.2 containing 301 rows and 13 columns is used for the Orange Data mining analysis. An additional column is added to the dataset, the employment outcome variable. This variable is computed on the Google Spreadsheet platform, categorizing each respondent's employment outcome as either "Successful" or "Unsuccessful" based on the values held in the previous 5 variables: Current Employment status, Time Elapsed before computing job, Persistence in Computing field, Highest Computing Employer Ranking, Highest Computing Salary Level. The values in the employment outcome section are computed following the same steps in section 3.3.2.2 but using spreadsheet formulas. Because women and blacks are the people groups under observation, this dataset is streamlined to just women and blacks, and rows with N/As (missing values) within the educational variables are excluded from the dataset. The dataset now contains 69 rows and 14 columns. This dataset can be found here: Orange Dataset Merged Variables with Employment Outcome.

In addition, another dataset is used that has more details about each respondent's educational history. This data is used to see the exact educational choices that have contributed to the respondents' employment outcome. For example: Within the "Orange Dataset Merged Variables with Employment Outcome" dataset, merged variables like "Traditional Degree" and "Alternative Degree" are present; while within this second dataset, the variables that were merged into "Traditional Degree", for instance, are individually represented. The merging process is outlined in section 3.2.2. The constituent variables of the "Traditional Degree" merged variable are: Associates' Degree, Bachelor's Degree, Master's Degree, and Doctoral Degree, for instance. This second dataset is a more detailed version of the "Orange Dataset Merged Variables with Employment Outcome" dataset. It contains the constituent variables of the merged educational variables in the "Orange Dataset Merged Variables with Employment Outcome" dataset. This second dataset consists of 69 rows and 20 columns and can be accessed here: Orange Dataset Full Variables with Employment Outcome

3.3.3.1 Preparing Data for Classification Analysis

In preparation for analysis with the Orange Data Mining tool, the two datasets (described above) are both split into two: a training dataset and a testing dataset. According to Liu & Cocea (2017), 70% of both datasets make up the training datasets while 30% of the datasets make up the testing datasets. The training datasets contain the first 48 rows of the Orange datasets while the testing datasets contain the remaining 21 rows of the datasets.

In addition, the Employment outcome variable is excluded from the test datasets because the trained classification algorithms will attempt to predict the employment outcomes of the respondents in the test datasets. Therefore, the “Orange Dataset Full Variables with Employment Outcome train” dataset retains all the 20 columns while the employment outcome variable is excluded from the “Orange Dataset Full Variables without Employment Outcome test” dataset, leaving only 19 variables within the testing dataset. Similarly, the “Orange Dataset Merged Variables with Employment Outcome train” dataset retains all the 14 columns while the employment outcome variable is excluded from the “Orange Dataset Merged Variables without Employment Outcome test” dataset, leaving only 13 variables within the testing dataset.

The “Orange Dataset Full Variables with Employment Outcome” dataset is split into: “Orange Dataset Full Variables with Employment Outcome train” and “Orange Dataset Full Variables without Employment Outcome test”. Similarly, the “Orange Dataset Merged Variables with Employment Outcome” dataset is split into “Orange Dataset Merged Variables with Employment Outcome train” and “Orange Dataset Merged Variables without Employment Outcome test”.

These training and testing datasets are used in the first phase of the predictive analysis with classification algorithms as specified in sections 3.3.8.1 and 4.1.5.1.

In the second phase of the predictive analysis with classification algorithms as specified in section 4.1.4.2, the original datasets: “Orange Dataset Full Variables with Employment Outcome” and “Orange Dataset Merged Variables with Employment Outcome” are stripped further of the 5 dependent variables that make up the employment outcome variable (the process of obtaining the employment outcome variable from these 5 dependent variables was described in section 3.3.2.2). The 5 variables: Current Employment status, Time Elapsed before computing job, Persistence in Computing field, Highest Computing Employer Ranking, Highest Computing Salary Level are excluded from the dataset. After excluding the 5 variables, these datasets are now named: "Orange Dataset Stripped Full Variables with Employment Outcome" and "Orange Dataset Stripped Merged Variables with Employment Outcome".

Similar to the datasets used in the first phase of Orange predictive analytics, the "Orange Dataset Stripped Full Variables with Employment Outcome" and "Orange Dataset Stripped Merged Variables with Employment Outcome" datasets are split into training and testing datasets, with the first 48 rows making up the training dataset and the last 21 rows making up the testing dataset. The employment outcome variables are also excluded from the testing datasets.

Therefore, the "Orange Dataset Stripped Full Variables with Employment Outcome train" dataset retains 15 columns while the employment outcome variable is excluded from the "Orange Dataset Stripped Full Variables without Employment Outcome test" dataset, leaving only 14 variables within the testing dataset. Similarly, the "Orange Dataset Stripped Merged Variables with Employment Outcome train" dataset retains 9 columns while the employment outcome variable is excluded from the "Orange Dataset Stripped Merged Variables without Employment Outcome test" dataset, leaving only 8 variables within the testing dataset.

The resulting training and testing datasets are: "Orange Dataset Stripped Full Variables with Employment Outcome train", "Orange Dataset Stripped Full Variables without Employment

Outcome test”, “Orange Dataset Stripped Merged Variables with Employment Outcome train”, and “Orange Dataset Stripped Merged Variables without Employment Outcome test”.

These training and testing datasets are used in the second phase of the predictive analysis with classification algorithms as specified in sections 3.3.8.2 and 4.1.5.2.

3.3.3.2 Preparing Data for Clustering

For the clustering data analysis, as specified in sections 3.3.7 and 4.1.4, the “Orange Dataset Stripped Full Variables with Employment Outcome” and “Orange Dataset Stripped Merged Variables with Employment Outcome” are used.

3.3.4 Conceptual Framework

The conceptual framework of this work shows a summary of the relationship between the independent and dependent variables which are relevant to the data analysis to be carried out. As shown in Figure 3.1 below, the hypothesis is that the educational pathways of women and blacks have an impact on their employment outcomes (either Successful or unsuccessful employment outcome). Hence, the independent variables consist of Race, Gender, Bachelors’ Degree and higher, Some College, Traditional Degree, Alternative Degree, Computing Degree, Rank1, Rank3, and Internship. These are the predictors for the dependent variable: Employment Outcome.

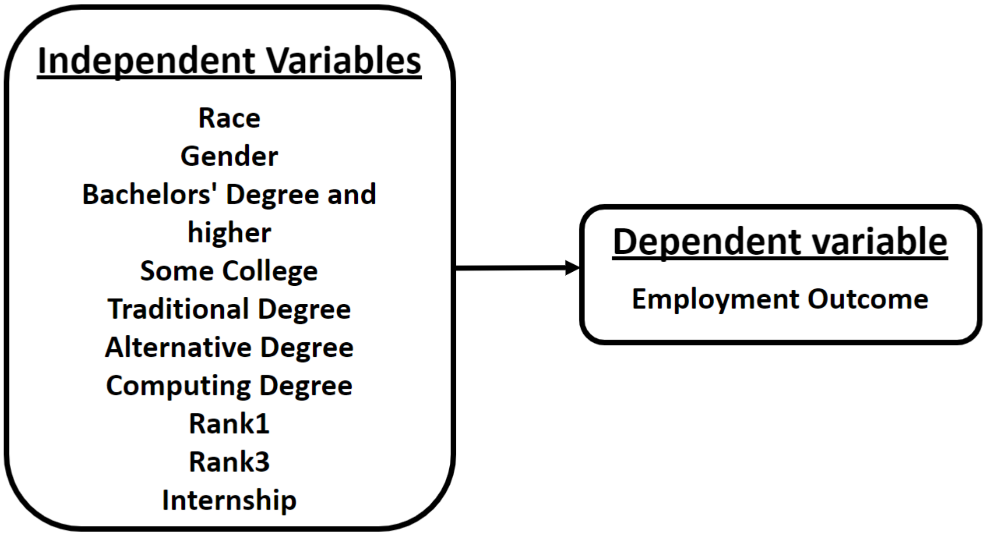

Figure 3.1 Conceptual Framework of Methodology

3.3.5 Univariate Analysis (in Stata) for extracting pathway themes for Females and Blacks

In order to answer the first research question which reads thus: "What are the common themes across educational pathway experiences that emerge from the analysis of computing professionals' data across racial and gender dimensions?", descriptive statistics are used. Univariate statistical analysis is employed to observe the various variables and to extract the current distribution of women and blacks across these variables. Since the objective of univariate analysis is to identify patterns within data by carrying out some descriptive statistics on the data (Hossain, 2019), univariate analysis would help to describe current patterns and themes that are in play regarding the educational pathways of women and blacks.

To answer the first research question, two-way tables (crosstabs) are used in Stata to show the relationship of Race and Gender with other variables in the dataset.

3.3.6 Logistic Regression and Predicted Probabilities (in Stata) for Predictive Data Analysis

In order to answer the second research question which reads thus: "Which of these common experiences result in successful long-term (greater than 3 years) employment outcomes in the technology sector for women and for blacks?", multivariate statistics (logistic regression and predicted probabilities) is used to measure the impact of the educational pathways of women and blacks on achieving successful employment outcomes.

3.3.6.1 Logistic Regression

Logistic regression is used to predict the relationship between the independent (predictor) variables and dependent (target) variable, as seen in the conceptual framework in Figure 3.1. Stata 17 is used for the logistic regression analyses. All the independent and dependent variables are recoded, a common data sample is found, and descriptive statistics is run resulting in a final sample size of 194 respondents, as described in Section 3.2.2.

Since little or no multicollinearity (high correlations among predictor variables) is required among the independent variables in logistic regression (StatisticsSolutions, 2021; Allison, 2012), the resulting model is checked for multicollinearity. After an initial run of the logistic regression model, the independent variables: Computing Degree, Bachelors' Degree and Higher, and Some College are omitted from the model due to multicollinearity. By the next run of the logistic regression model, those 3 variables are manually excluded from the model so as to increase the reliability of the regression results.

The logistic regression is able to clearly show the individual relationships between each predictor variable and the target variable.

3.3.6.2 Predicted Probabilities

Within Stata, predicted probabilities are employed in order to predict the target variable from any combination of the predictor variables. These predicted probabilities are based on the results of the logistic regression model. Predicted probabilities goes beyond just clarifying the impact of the individual predictor variables on the target variable; instead, it can predict the target variable from any combination of attributes of the independent variables.

This functionality is useful for the exploration of the second research question that seeks to paint a picture of the pathways or educational characteristics of women and blacks that have a successful outcome. Predicted probabilities are used, in this research, to describe the educational factors that result in a high probability of successful employment outcomes for women and blacks, and educational factors that have less probability of success.

3.3.7 Clustering (in Orange) for Descriptive Data Analysis

To support the results from the Predicted Probabilities analysis in the previous section, the k-means clustering algorithm in Orange Data Mining is employed. Even though k-means clustering is usually used for unlabeled data, which is continuous in nature, it is used in this research to group together people with similar employment outcomes, and to determine what characterizes each group.

In the clustering experiment, k = 2. That is, the data is grouped into two clusters: the "Successful" and "Unsuccessful" attributes of the Employment Outcome variable, and the most recurring educational attributes/pathways of the successful data respondents are identified. The k-means algorithm is run 10 times, from random initial positions, and the result with the lowest within-cluster sum of squares will be used. The maximum number of iterations within each run of the algorithm is set to 300.

Figure 3.2 shows the structure of the clustering model in Orange. File (1) contains the full dataset to be used in this data analysis. The distribution of the data across the two clusters, Successful and Unsuccessful, is described in section 4.1.4.

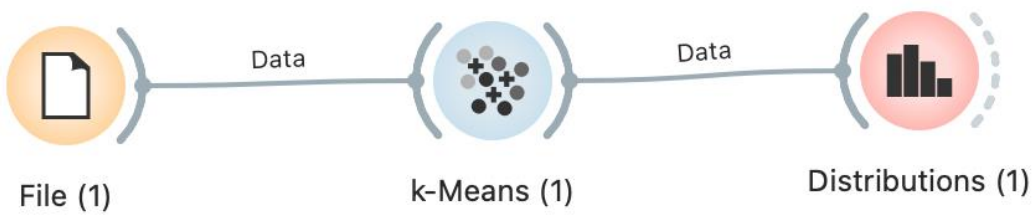

Figure 3.2 Framework of Clustering Analysis Model

This data analysis technique is useful to investigate the second research question that seeks to paint a picture of the pathways or educational characteristics of women and blacks that have a successful outcome.

3.3.8 Classification Algorithms (in Orange) for Predictive Data Analysis

This data analysis methodology is employed, in addition to the Logistic Regression and Predicted Probabilities analysis in Stata, to carry out predictive analysis using several Machine Learning (ML) algorithms within the Orange Data Mining tool. Within the Orange Data Mining tool, the

Naïve Bayes, Logistic Regression, Neural Network, and Random Forest predictive techniques are employed.

The Naïve Bayes classifier learns a Naïve Bayesian model from the data and outputs a Naïve Bayes learning algorithm and a trained model. The Naïve Bayes classifier in Orange preprocesses the input data by removing empty columns and discretizing numeric values to 4 bins with equal frequency.

The Logistic Regression classifier uses ridge (L2) regularization with the cost strength of the regularization parameter set to 1. The Logistic Regression classifier in Orange preprocesses the input data by removing instances with unknown target values, converting categorical variables to continuous variables using one-hot-encoding, removing empty columns, and imputing missing values with mean values.

The Neural Network classifier uses sklearn's Multi-layer perceptron algorithm. Within the Neural Network classifier, there are 100 neurons in the hidden layer. The hidden layer uses the rectified linear unit activation function. The Adam solver, a stochastic gradient-based optimizer is used for weight optimization. For regularization, the L2 parameter, alpha, is used, and is set to 0.00010. This parameter is useful to prevent overfitting in the neural network machine learning model. The maximum number of iterations of the classifier's learning process is set to 200. The Neural Network classifier in Orange preprocesses the input data by removing instances with unknown target values, converting categorical variables to continuous variables using one-hot-encoding, removing empty columns, imputing missing values with mean values, and normalizing the data by centering to mean and scaling to standard deviation of 1.

Finally, the Random Forest classifier builds a set of decision trees, 10 trees in this context. The number of attributes that are considered at each node (before a split occurs) is defined by the square root of the total number of attributes in the data. The size of the smallest subset that can be split is set to 5. The Random Forest classifier in Orange preprocesses the input data by removing instances with unknown target values, converting categorical variables to continuous variables using one-hot-encoding, removing empty columns, and imputing missing values with mean values.

The sampling details of the classification procedures include the k-fold cross-validation where k=5, random sampling repeated 10 times with the training set size of 70% and testing set size of 30%. The prediction results are generated for the "Successful" target class. That is, the measures of the ability of the classification algorithms to predict a "Successful" employment outcome are reported.

The structure of the classification model, using all 4 classifiers, in Orange is shown in Figure 3.3, where File (1) holds the training dataset and File holds the testing dataset. The prediction scores are discussed in section 4.1.5.

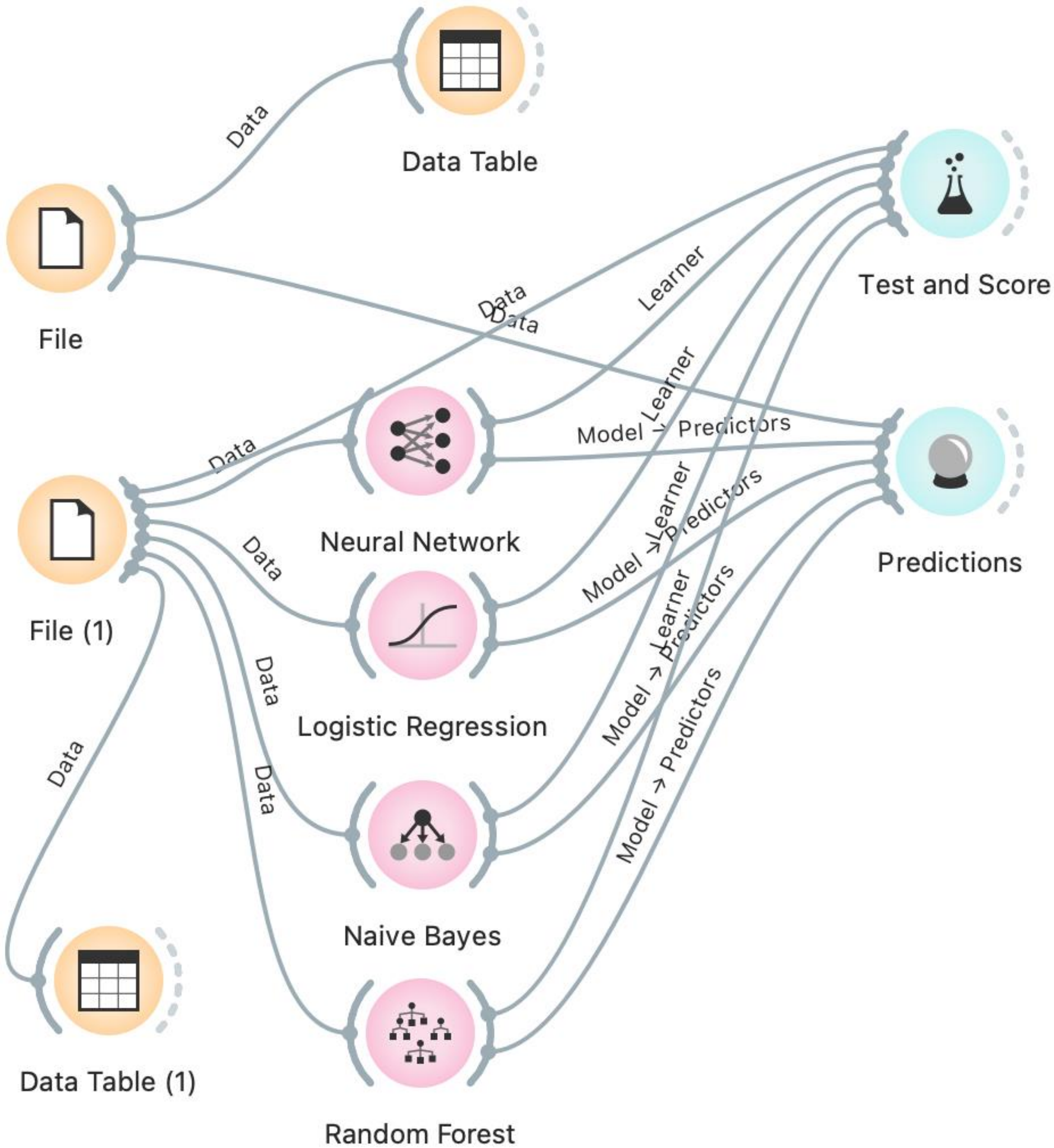

Figure 3.3 Framework of Classification Analysis Model

The classification analysis aims to look ahead to determine the applicability of the data analysis results to future data. The classification algorithms aim to learn the relationship between the educational history of respondents and their employment outcome and then apply what has been learnt to determine the employment outcome of previously unseen data. This data analysis method measures how accurately the relationships established by previous data analysis methods can be generalized (applied to the general population, beyond the specific dataset used in this study).

Because of the very limited size of the training datasets (48 records), it is expected that the machine learning (ML) algorithms will not have enough data to build accurate models. Therefore, the resulting predictive accuracy of the ML models can be improved by feeding the algorithms with more training data, which is not readily available within this research.

First Phase of Predictive Analysis using Classification Algorithms

In the first phase of predictive analysis with classification algorithms, both the educational and employment details of respondents are used to predict their employment outcome.

3.3.8.2 Second Phase of Predictive Analysis using Classification Algorithms

In the second phase of predictive analysis with classification algorithms, only the educational details of respondents are used to predict their employment outcome. The employment details of respondents were excluded from this phase because of the high correlation between the employment details and the employment outcome variable.

3.3.9 Data Analysis Results, the Current State of Computing Education, and Next Steps

In order to answer the third research question which asks: "How do the findings of this study inform national investment in broadening participation efforts that seek to increase racial and gender diversity in the computing workforce?", the data analysis results will be discussed and possible applications to the national "broadening participation" efforts will be proposed, based on the current state of operations at different computing educational institutions.

CHAPTER IV

DATA ANALYSIS AND DISCUSSION OF RESULTS

4.1 SOME PRELIMINARY RESULTS

4.1.1 Univariate Analysis Results

After examining several pairs of variables using two-way tables in Stata, where each pair includes either a Race or Gender variable, here are the themes that are observed within the data.

- 75% of black people do not have an internship.
- 69.8% of women do not have an internship.
- 93.2% of blacks have a traditional degree.
- All women (100%) in our dataset have a traditional degree.
- 56.8% of blacks do not have an alternative degree.
- 55.8% of women do not have an alternative degree.
- 66.7% of blacks attended a Rank1 institution.
- 83.3% of women attended a Rank1 institution.
- 73.8% of blacks did not attend a Rank2 institution.
- 83.3% of women did not attend a Rank2 institution.
- 92.9% of blacks did not attend a Rank3 institution.
- 100% of women did not attend a Rank3 institution.
- 62.2% of blacks have a successful employment outcome.

- 67.6% of women have a successful employment outcome.

We can see from the above percentages that majority of black people did not do an internship, have a traditional degree, do not have an alternative degree, attended a Rank1 institution, and have a successful employment outcome. Along the gender lines, we see that majority of females did not do an internship, have a traditional degree, do not have an alternative degree, attended a Rank1 institution, and have a successful employment outcome. The education trends appear similar for both blacks and females: possession of a traditional degree, attendance of a Rank1 institution, no internship. Do these educational choices guarantee successful computing employment outcomes among females and blacks? Logistic regression and predicted probabilities were run on the dataset to show what educational choices are good predictors for successful employment outcomes.

4.1.2 Logistic Regression Results

The final logistic regression model consists of the following independent variables: Race, Gender, Traditional Degree, Alternative Degree, Rank1, Rank3, and Internship, with Employment Outcome as the dependent variable. This model was then run in Stata and the table below displays the results of the regression.

Table 4.1 Logistic Regression Results

Variables	**B (Coefficient)**	**Standard Error**	**Odds Ratio**	**P**
Race	-0.631	0.423	0.532	0.136
Gender	-0.366	0.448	0.694	0.415
Traditional Degree	1.665	1.521	5.288	0.274
Alternative Degree	0.938	0.378	2.556	0.013
Rank1	0.874	0.396	2.397	0.027
Rank3	1.171	0.975	3.225	0.230
Internship	0.555	0.420	1.742	0.186
Constant	-1.645	1.600	0.193	0.304
Pseudo R2	0.065			
Overall significance of model	0.0357			
Number of observations	197			

The Odds ratio measures the ratio of the odds (of predicting the dependent variable) for the variable attribute that is set to 1 (the group under observation), in relation to the odds of the variable attribute that is set to 0 (the reference group).

From our results, for instance, the odds are about 0.5 to 1 that a black person will have a successful computing employment outcome compared to non-blacks. That is, there is a lesser likelihood of blacks achieving successful computing employment, when compared to non-blacks. Similarly, the odds are about 5 to 1 that an individual with a traditional degree will have a successful computing employment outcome compared to an individual without a traditional degree.

As the above table shows, having a traditional degree increases a person's likelihood for obtaining successful employment outcomes by 5 times of the likelihood of an individual who does not possess a traditional degree. Having an alternative degree also gives an individual 3 times more likelihood of obtaining successful computing employment than an individual who does not possess an alternative degree. Similarly, an individual who does an internship is almost 2 times more likely to achieve a successful computing employment than someone who does not do an internship.

The Coefficient measures the amount of change expected in the log-odds when there is a unit change in the independent variable, while all other variables remain unchanged. The coefficient also shows a similar relationship between the independent variable and dependent variable as can be seen in the Odds ratio variable. A negative coefficient shows a lesser likelihood of the group under observation to achieve the target than the reference group. For the Race variable, Blacks make up the group under observation while non-blacks are the reference group. On the other hand, a positive coefficient indicates a higher likelihood of the group under observation to achieve the target than the reference group. Therefore, our regression table shows, for instance, that blacks are expected to have 0.631 fewer log-odds of achieving a successful employment outcome than non-blacks, and females are expected to have 0.366 fewer log-odds of achieving a successful employment outcome than males. Similarly, people with a traditional degree are expected to have 1.665 more log-odds of achieving successful employment outcome than those who do not possess a traditional degree.

The Standard error shows how much each variable's coefficient differs from 0, measuring the statistical accuracy of the coefficient of each variable (Google, 2021; UCLA, 2021).
The "Constant" variable represents the value of the dependent variable when all the independent variables are set to zero. In our table, when a white male does not possess a traditional degree or an alternative degree and does not attend either a Rank 1 or Rank 3 institution, and does not do an internship, he has odds of 0.2 to 1 of attaining a successful employment outcome.

As seen in the table, a total of 197 data points were analyzed. Pseudo R2 represents the relevance of the independent variables in this model to predicting the dependent variable. The pseudo R2 of 0.065 shows that 6.5% of obtaining a successful employment outcome is influenced (or can be predicted) by the independent variables used in this model. This shows that the pool of independent variables needs to be expanded. The overall regression model is statistically significant because its p-value equals 0.0357 which is less than 0.05. This means that the results of our overall model (that is, the ability of all the independent variables to predict the dependent variable) falls within a 95% confidence interval.

On the other hand, only two of the independent variables possess statistical significance (when standing alone) with respect to their ability to predict the dependent variable. It can be seen from Table 2 that Alternative degree (p-value: 0.013) and Rank1 (p-value: 0.027) are the only independent variables that are statistically significant in their individual relationships with the dependent variable.

4.1.3 Predicted Probabilities Results

4.1.3.1 General Results

After running predicted probabilities in Stata, the results in Figure 4.1 showed that men and non-blacks have a higher probability of attaining more successful computing employment outcomes than women and blacks. This supports the claim of the literature that women and blacks are underrepresented in computing employment. Since both women and blacks are at a disadvantage (compared to men and non-blacks), it follows that black women would be the least probable to attain successful computing employment as seen in Figure 4.2.

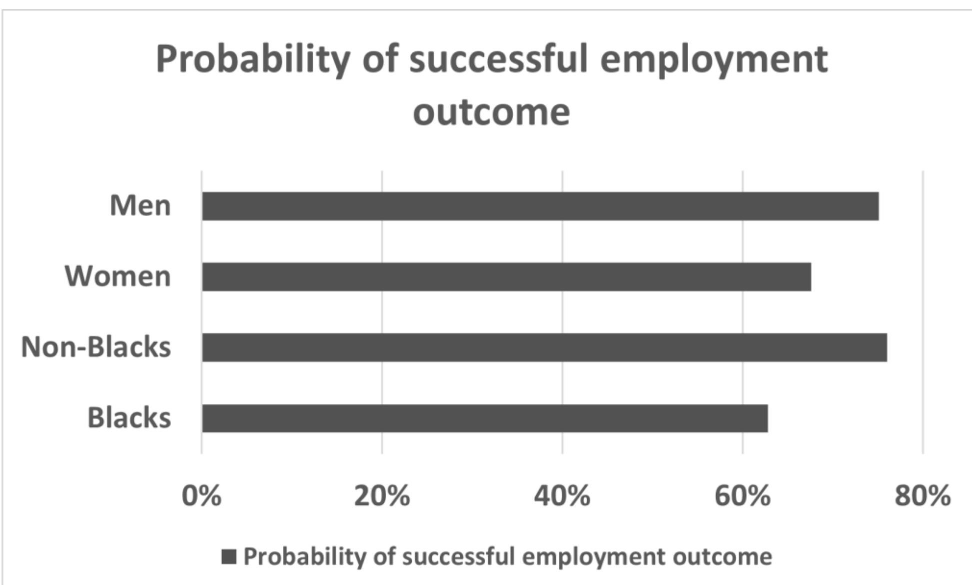

Figure 4.1 Probability of successful employment outcome (by Race and Gender – separately)

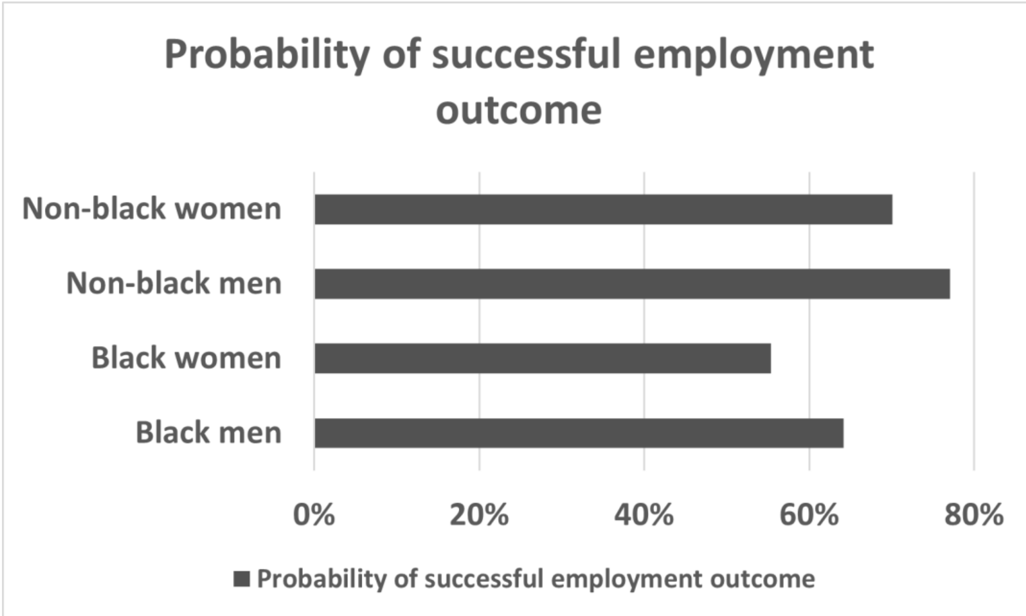

Figure 4.2 Probability of successful employment outcome (by Race and Gender – together)

4.1.3.2 Results along the lines of Degree Type

As seen in Figure 4.3, men and women with any form of alternative degree (either alone or in addition to a traditional degree) are shown to be more likely to achieve successful employment outcomes than those with only a traditional degree. In this context, an alternative degree consists of a computing certification, coding bootcamp, or some other computing degree (apart from the traditional degree). The traditional degree, on the other hand, refers to an associate's degree, bachelor's degree, master's degree, or a doctoral degree.

Similarly, blacks and non-blacks with alternative degrees have a higher chance of successful computing employment than those without an alternative degree, as seen in Figure 4.4.

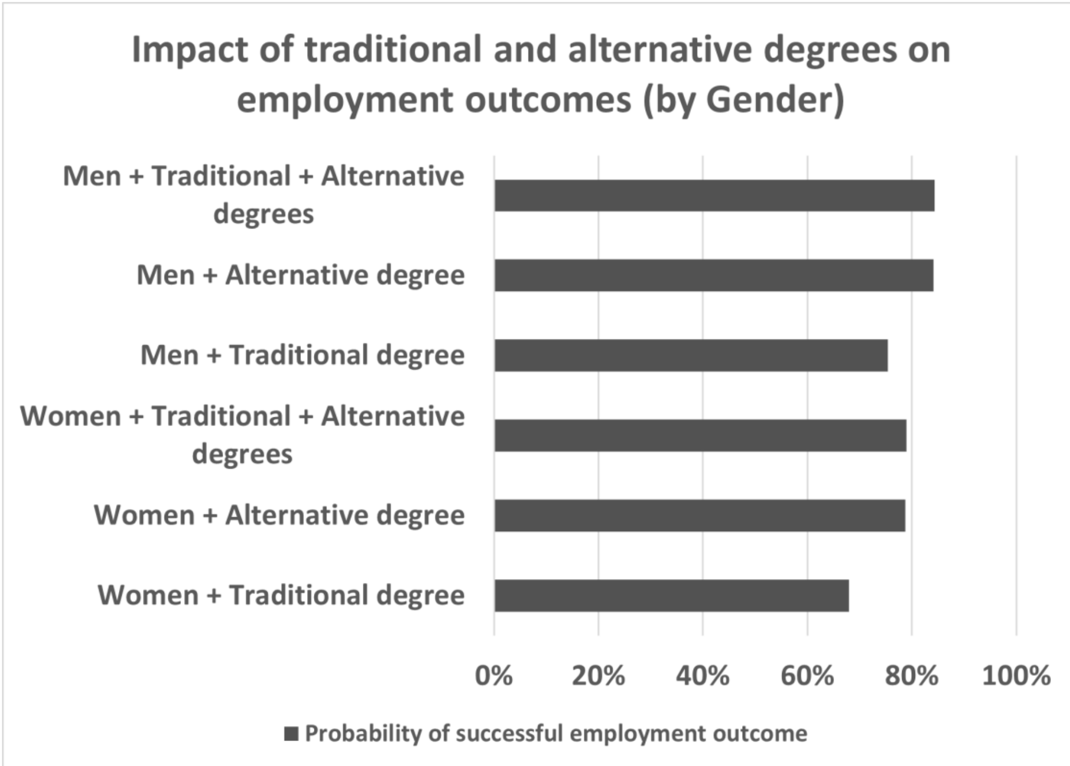

Figure 4.3 Impact of traditional and alternative degrees on employment outcomes (by Gender)

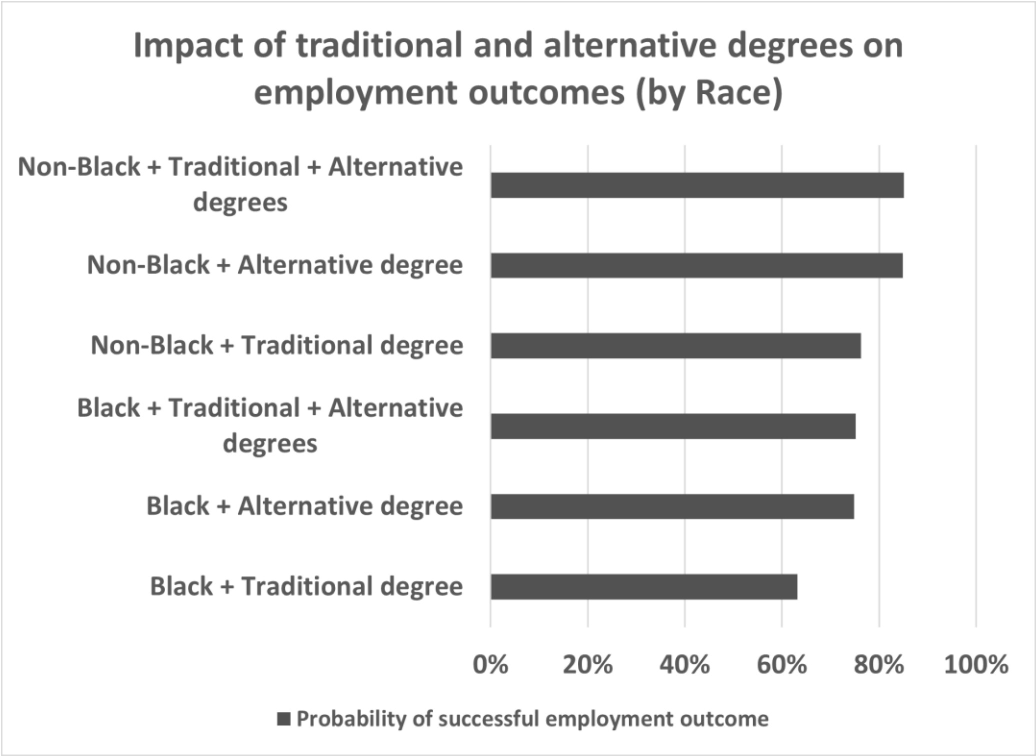

Figure 4.4 Impact of traditional and alternative degrees on employment outcomes (by Race)

4.1.3.3 Results along the lines of Institution Ranking

As seen in Figure 4.5, blacks who attend a Rank 3 institution (that is, Associate degree-granting colleges) have a higher probability of attaining successful computing employment outcomes than those who attend a Rank 1 institution (R1, R2, and Doctoral/Professional universities). This is contrary to what the results show about non-blacks, as also seen in Figure 4.5.

Among the blacks, those who have an alternative degree and an education from a Rank 3 institution possess the highest chance of achieving successful computing employment outcomes. This can be seen in Figure 4.6.

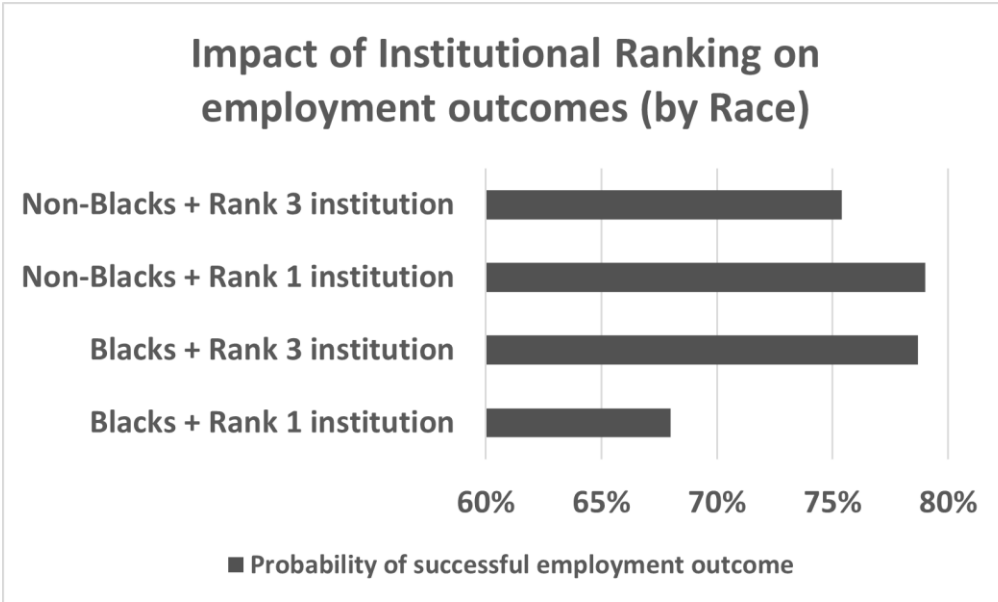

Figure 4.5 Impact of Institutional Ranking on employment outcomes (by Race)

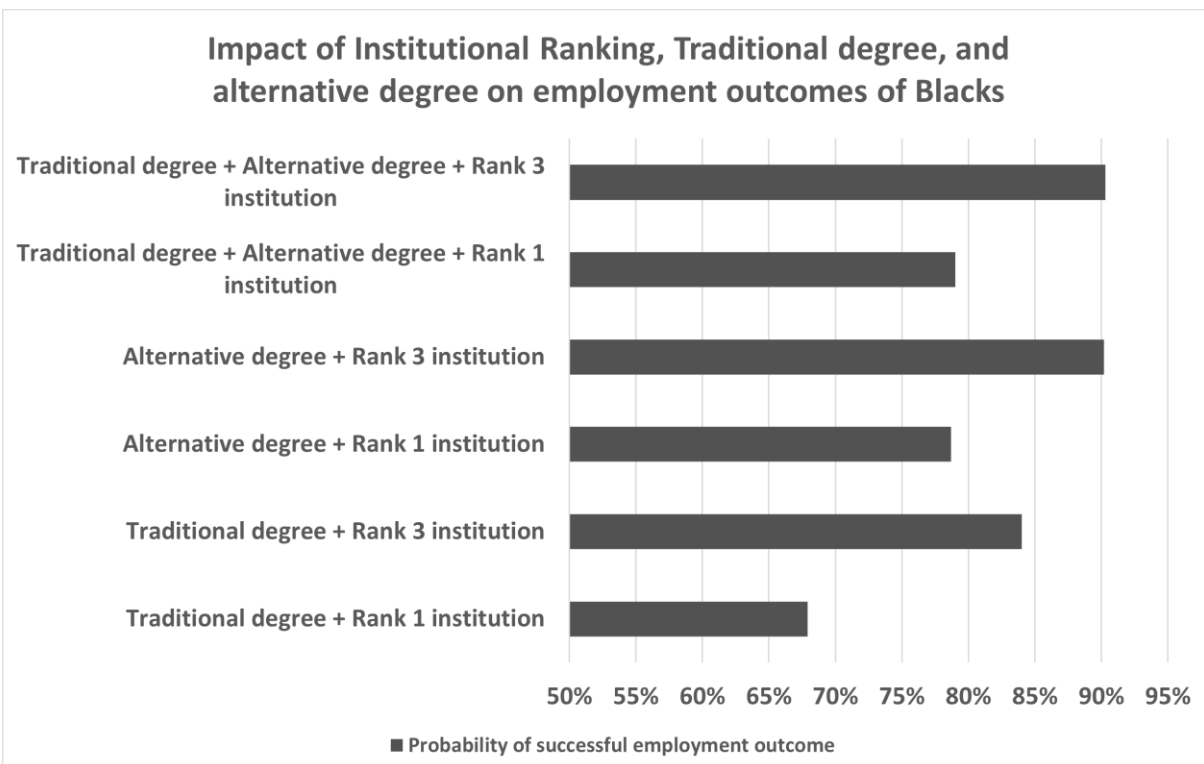

Figure 4.6 Impact of Institutional Ranking, Traditional Degree, and Alternative Degree on employment outcomes of Blacks

Slightly different trends are seen along gender lines. As Figure 4.7 shows, both men and women experience a higher probability of successful employment outcomes when they attend a Rank 3 institution, compared to when they attend a Rank 1 institution.

Figure 4.8 also shows that, for women, possessing an alternative degree and attending a Rank 3 institution yields the highest probability for attaining successful employment outcomes. This is similar to the findings for black people.

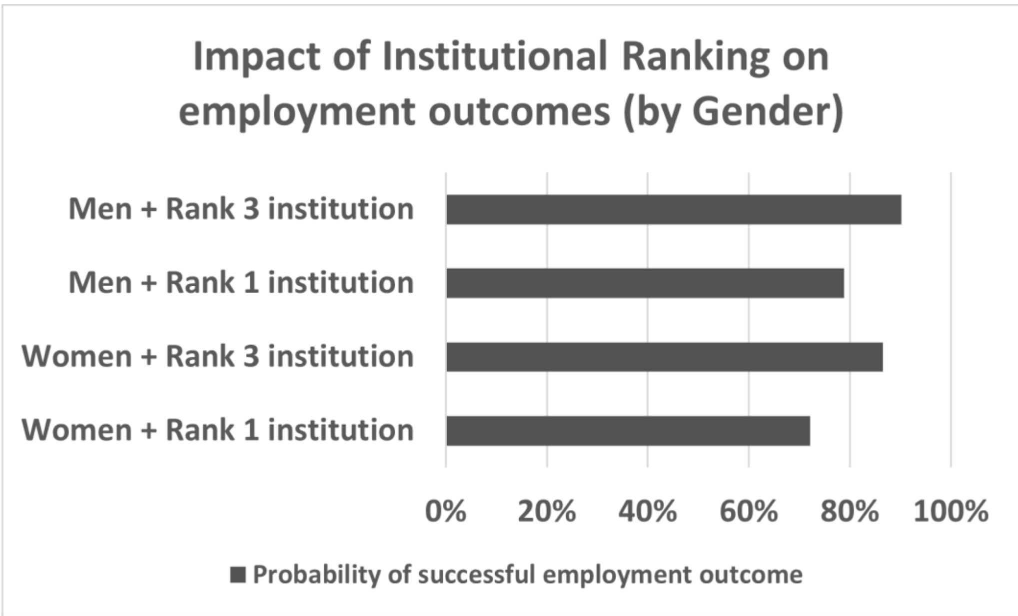

Figure 4.7 Impact of Institutional Ranking on employment outcomes (by Gender)

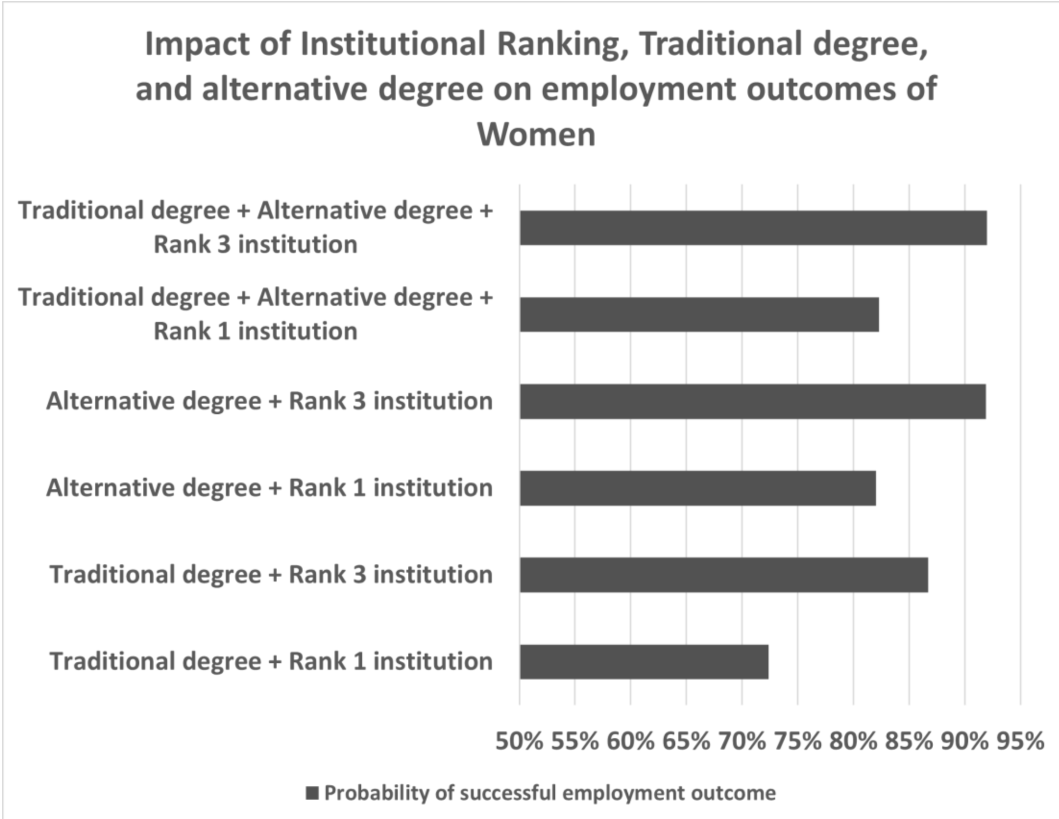

Figure 4.8 Impact of Institutional Ranking, Traditional Degree, and Alternative Degree on employment outcomes of Women

4.1.3.4 Results along the lines of Internship

Men and Non-blacks with computing internships have recorded a higher probability of achieving successful computing employment outcomes than women and blacks who also did computing internships, as seen in Figure 4.9.

Blacks and Women who possess an alternative degree and who also did a computing internship have been found to be more likely to attain successful computing employment outcomes than those who had a traditional degree and computing internship, as seen in Figures 4.10 and 4.11.

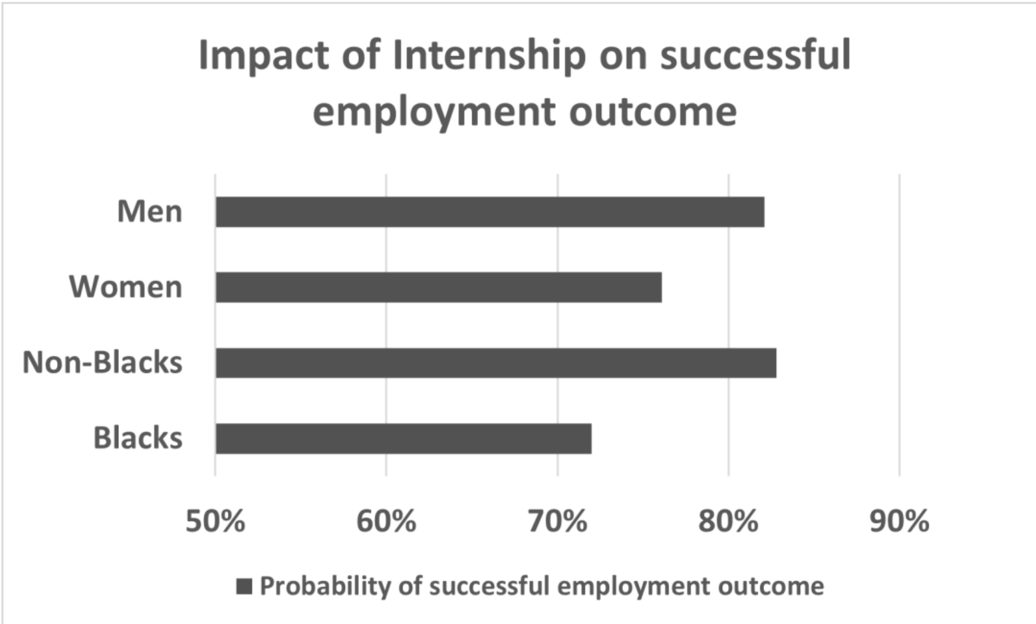

Figure 4.9 Impact of Internship on successful employment outcomes (by Race and Gender)

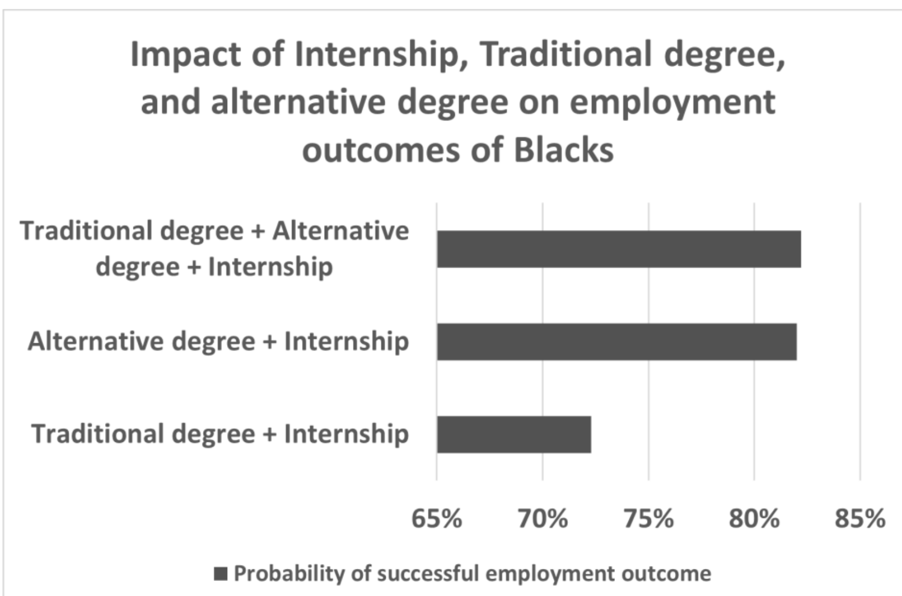

Figure 4.10 Impact of Internship, Traditional degree, and alternative degree on employment outcomes of Blacks

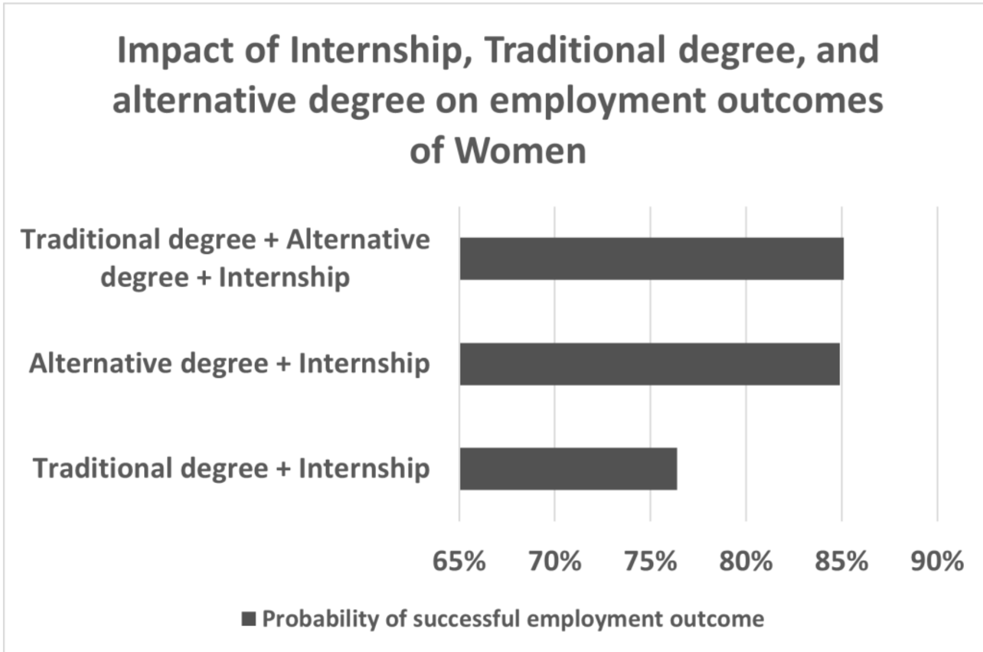

Figure 4.11 Impact of Internship, Traditional degree, and alternative degree on employment outcomes of Women

4.1.3.5 Results along the lines of individual Independent Variables

Examining the individual impact of each independent variable on the employment outcomes of blacks and women shows that the single independent variable that provides women and blacks with the highest probability of a successful employment outcome is the Rank 3 institution variable. Hence, if a woman or a black person attends a Rank 3 institution, they are more than 80% likely to achieve successful computing employment outcomes. This is shown in Figures 4.12 and 4.13.

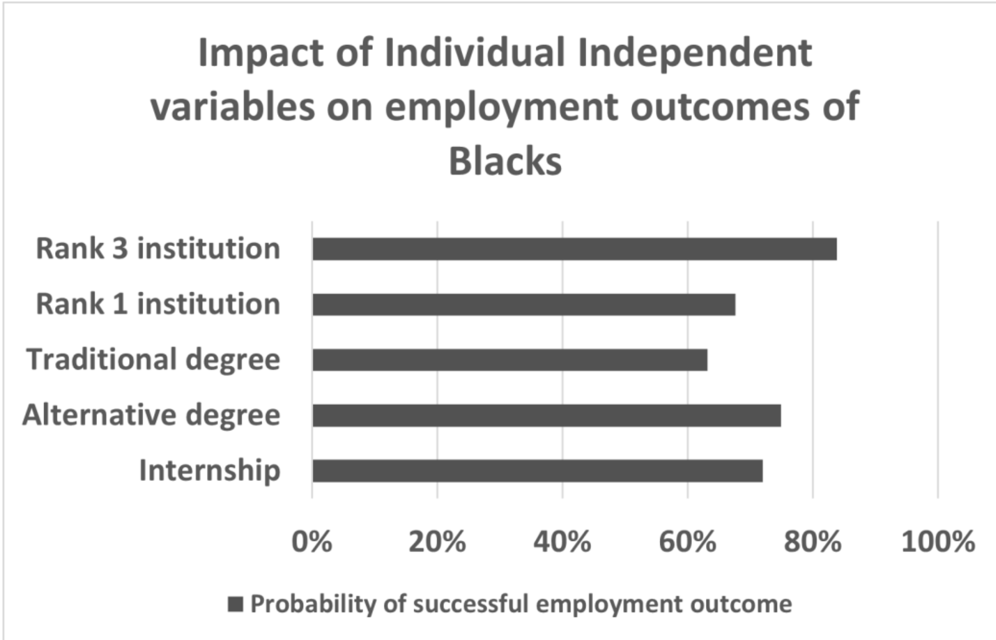

Figure 4.12 Impact of individual independent variables on employment outcomes of Blacks

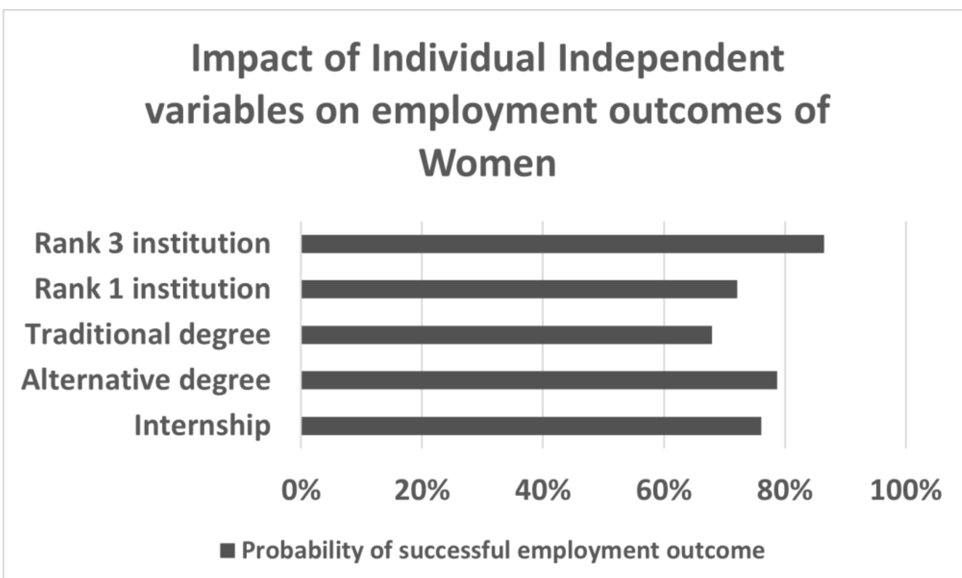

Figure 4.13 Impact of individual independent variables on employment outcomes of Women

4.1.3.6 Results along the lines of the combination of all Independent Variables

Putting all the independent variables together and measuring their combined impact on the probability of successful computing employment outcomes for women and blacks, it was found that women and blacks who have the highest probability of attaining successful employment outcomes are those who did an internship, possess both a traditional degree and an alternative degree, and attended a Rank 3 institution. This is pictured in Figures 4.14 and 4.15.

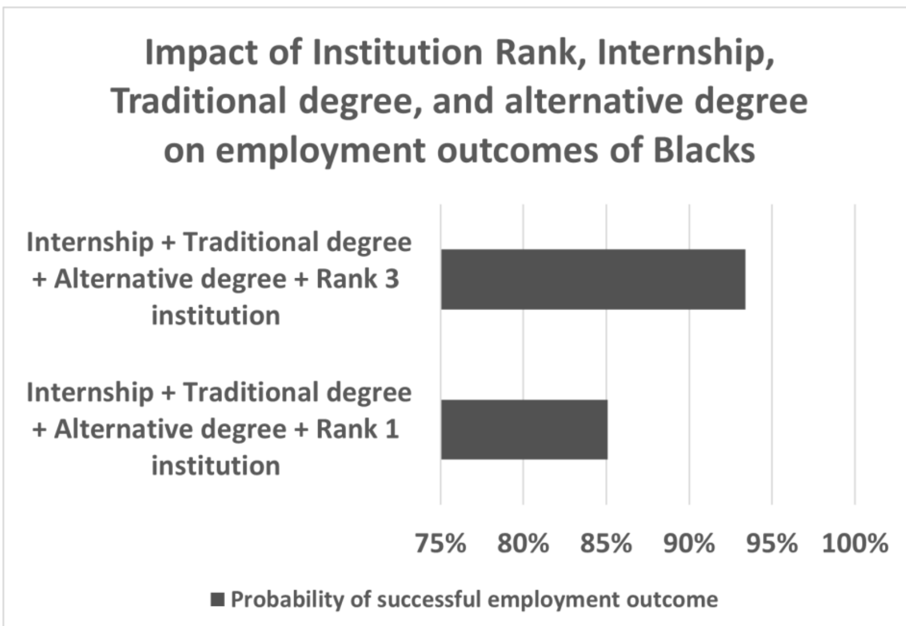

Figure 4.14 Impact of Institution Rank, Internship, Traditional Degree, and Alternative degree on the employment outcomes of Blacks

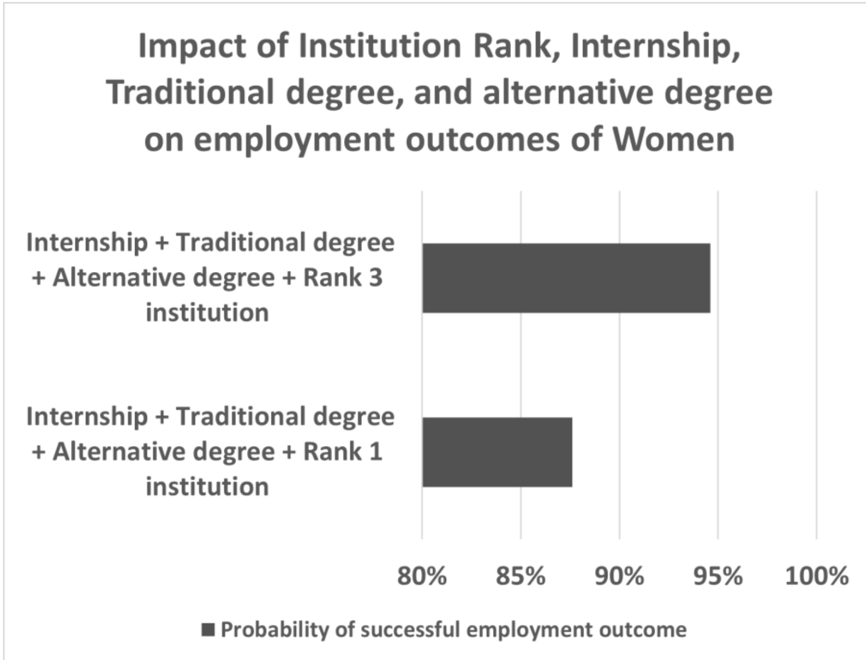

Figure 4.15 Impact of Institution Rank, Internship, Traditional Degree, and Alternative degree on the employment outcomes of Women

4.1.3.7 Pathways that produce the highest (and lowest) probabilities of success

From the analysis results, there are a few combinations of independent variables that result in a likelihood of employment success of 90% and above, for blacks and women:

- Probability of success for a black person who has a traditional degree, an alternative degree, and a Rank 3 institution education: **90.3%**
- Probability of success for a woman who has a traditional degree, an alternative degree, and a Rank 3 institution education: **92%**
- Probability of success for a black person with an internship, a traditional degree, an alternative degree, and a Rank 3 institution education: **93.4%**

- Probability of success for a <u>woman</u> with an <u>internship</u>, a <u>traditional degree</u>, an <u>alternative degree</u>, and a <u>Rank 3 institution</u> education: **94.6%**

The probabilities above are the highest probabilities of success for blacks and women as produced by the data analysis, and it shows us the combination of variables (or educational pathway choices) that are predicted to yield the highest probability of successful computing employment outcomes for blacks and women.

Based on this analysis, it is predicted that a black person who does an internship, possesses a traditional degree, possesses an alternative degree, and also attends a Rank3 institution has the highest likelihood to obtain successful computing employment outcomes. This also applies to women. Attending a Rank3 institution (that is, Associate degree-granting institution), possessing a traditional degree (which ranges from an associates' degree to a doctoral degree), possessing an alternative degree (which includes certifications, coding bootcamp degree, or other degree), and doing an internship is the pathway that is predicted to yield the highest probability of a successful employment outcome for blacks and women. Interestingly, this is not the pathway that the majority of blacks and women have taken, as seen in Section 4.1.1 where the majority of women and blacks possess a traditional degree, attended a Rank1 institution, and did not have an internship.

On the other hand, there are a few combinations of independent variables that result in a likelihood of employment success of less than 70%, for blacks and women:

- Probability of success for a <u>black</u> person with an <u>internship</u>, an <u>alternative degree</u>, and a <u>Rank 1 institution</u> education = **51.9%**
- Probability of success for a <u>woman</u> with an <u>internship</u>, an <u>alternative degree</u>, and a <u>Rank 1 institution</u> education = **57.2%**

- Probability of success for a black person with a traditional degree = **63.2%**
- Probability of success for a black person with a Rank1 institution education = **67.6%**
- Probability of success for a woman with a traditional degree = **68%**

The above list shows the combination of independent variables that result in the least probability of attaining successful computing employment outcomes, for blacks and women. A black person or a woman who possesses only a traditional degree or who only attends a Rank 1 institution (R1, R2, and Doctoral/Professional universities) or who attends a Rank 1 institution in addition to getting an alternative degree and an internship have been predicted as experiencing the least likelihood for computing employment success.

A common pathway thread that results in the least successful outcomes for blacks and females include either attending only a Rank 1 institution or having only a traditional degree. Interestingly, this is the pathway that the majority of women and black people have taken, as seen in Section 4.1.1 where majority of women and blacks possess a traditional degree, attended a Rank1 institution, and did not have an internship. This provides an explanation as to why women and blacks have not experienced a lot of successful computing employment outcomes.

4.1.4 Clustering Data Analysis Results

In this results section, each educational history variable is studied, examining the distribution of the employment outcomes of women and blacks who took that educational pathway. The charts below show the probability of a successful (or unsuccessful) employment outcome depending on whether (or not) a woman or black person chooses to take a particular educational pathway.

Figure 4.16 shows that women and blacks who take the route of getting an Associate's degree have a higher probability of an unsuccessful employment outcome than those who do not get an Associate's degree. On the other hand, Figure 4.17 shows that having a bachelor's degree is a predictor of a successful employment outcome in the computing industry while a lack of a bachelor's degree is a predictor of unsuccessful employment outcomes. Similarly, possessing a masters' degree is a predictor of a successful employment outcome while a lack of a masters' degree is a predictor of unsuccessful employment outcomes, as seen in Figure 4.18. Possessing a doctoral degree results in a higher chance of being successful than being unsuccessful, whereas not possessing a doctoral degree produces an equal probability of successful and unsuccessful employment outcomes. This can be seen in Figure 4.19.

The probability of attaining a successful outcome given the highest degree attained is shown in Figure 4.20, where having a high school degree (or lower) as the highest degree attained leads to a 100% probability of an unsuccessful outcome, having some college (that is, a 2-year college degree or less than a 4-year college degree) as the highest degree attained results in a 100% probability of a successful outcome, while having a bachelor's degree or higher (that is, bachelor's degree, masters' degree, or doctoral degree) as the highest degree attained yields higher probability of successful employment outcome than unsuccessful employment outcomes.

Figure 4.21 shows the impact of having a computing degree on attaining computing employment. Possessing a computing degree results in a high probability of successful employment outcome while not possessing a computing degree results in an even higher probability of an unsuccessful employment outcome.

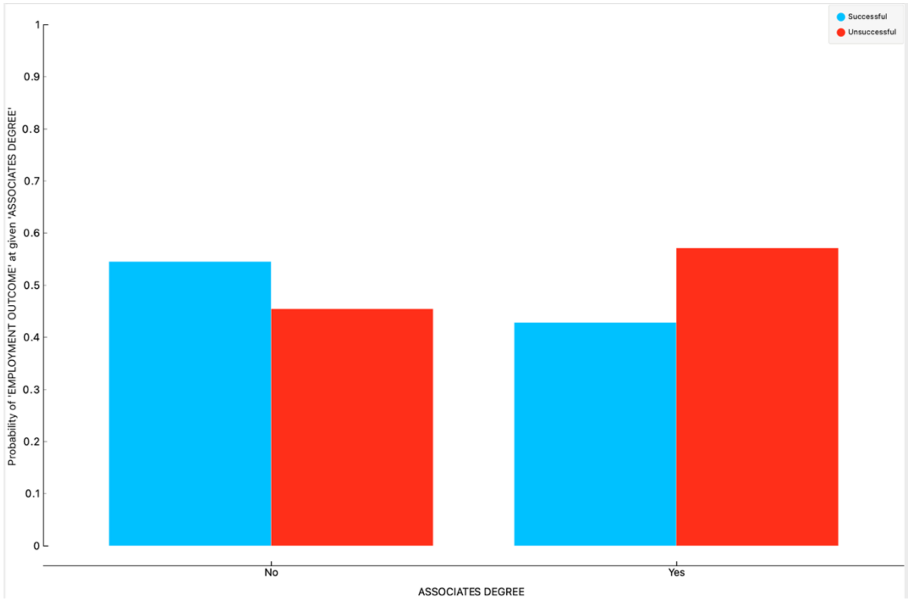

Figure 4.16 The probability of a successful (or unsuccessful outcome) given the possession of an Associate's degree

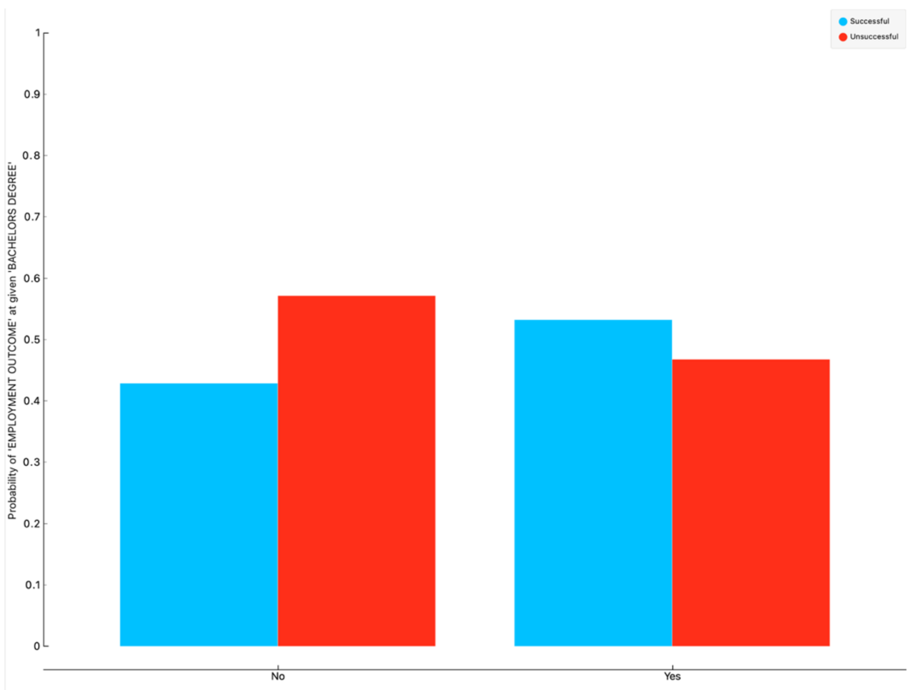

Figure 4.17 The probability of a successful (or unsuccessful outcome) given the possession of a Bachelor's degree

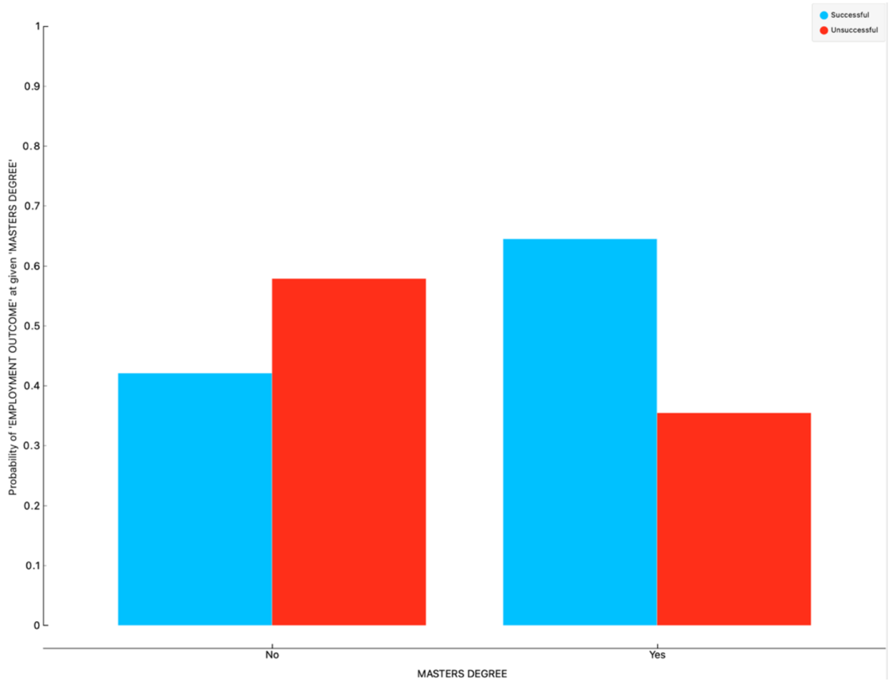

Figure 4.18 The probability of a successful (or unsuccessful outcome) given the possession of a Masters' degree

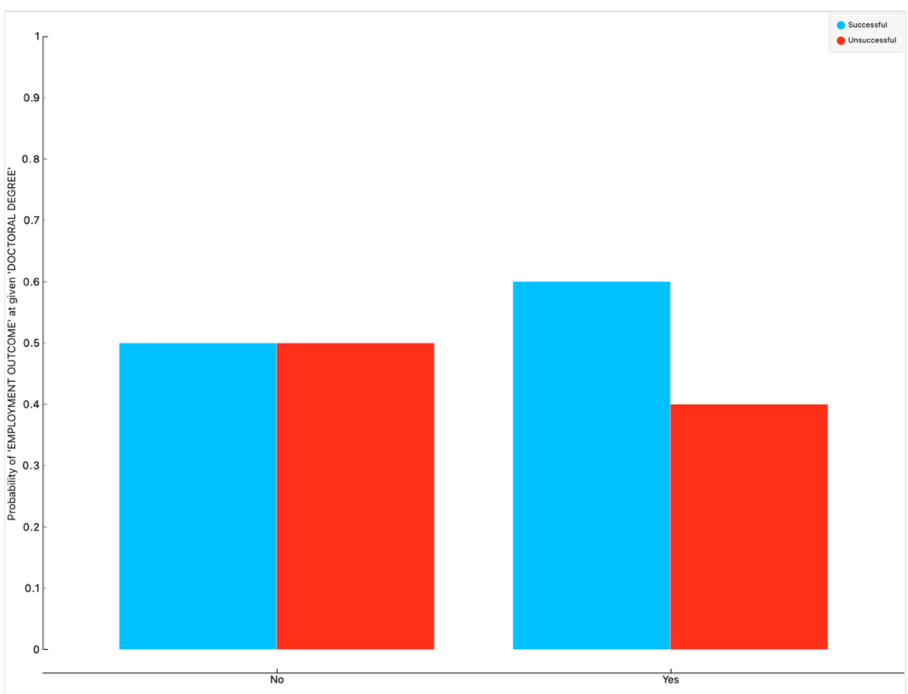

Figure 4.19 The probability of a successful (or unsuccessful outcome) given the possession of a Doctoral degree

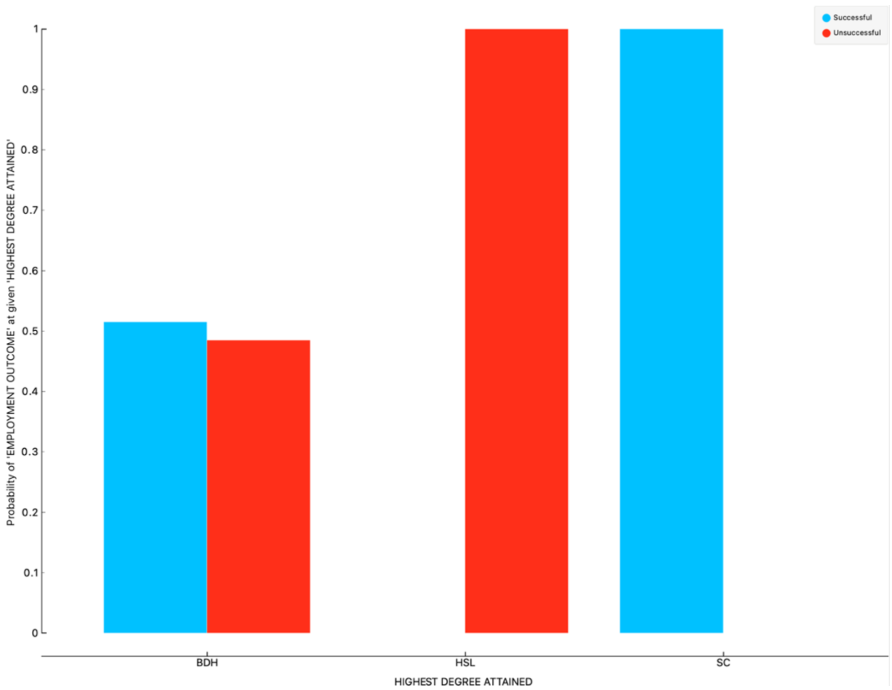

Figure 4.20 The probability of a successful (or unsuccessful outcome) given the highest degree attained

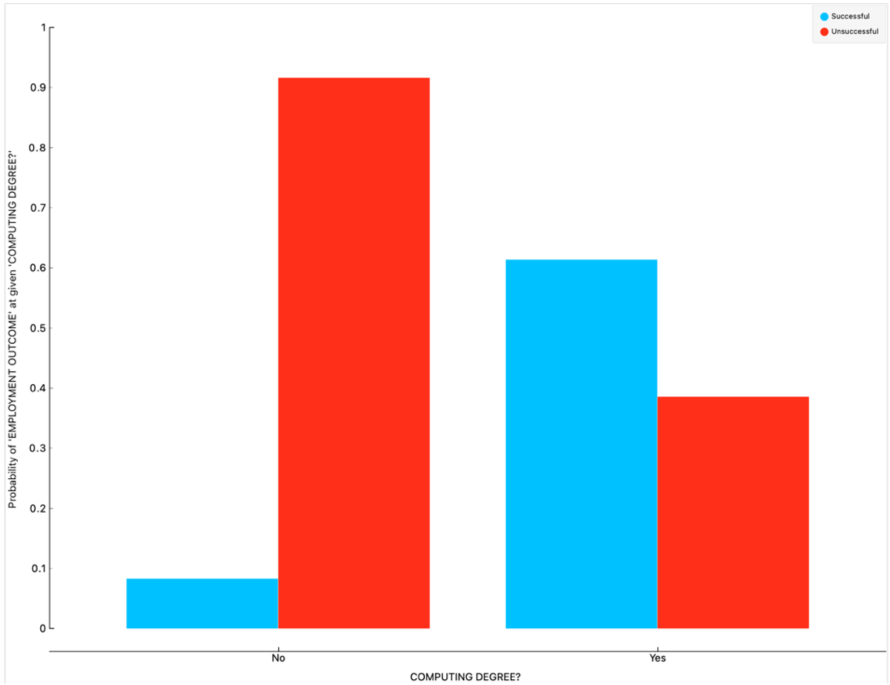

Figure 4.21 The probability of a successful (or unsuccessful outcome) given the possession of a computing degree

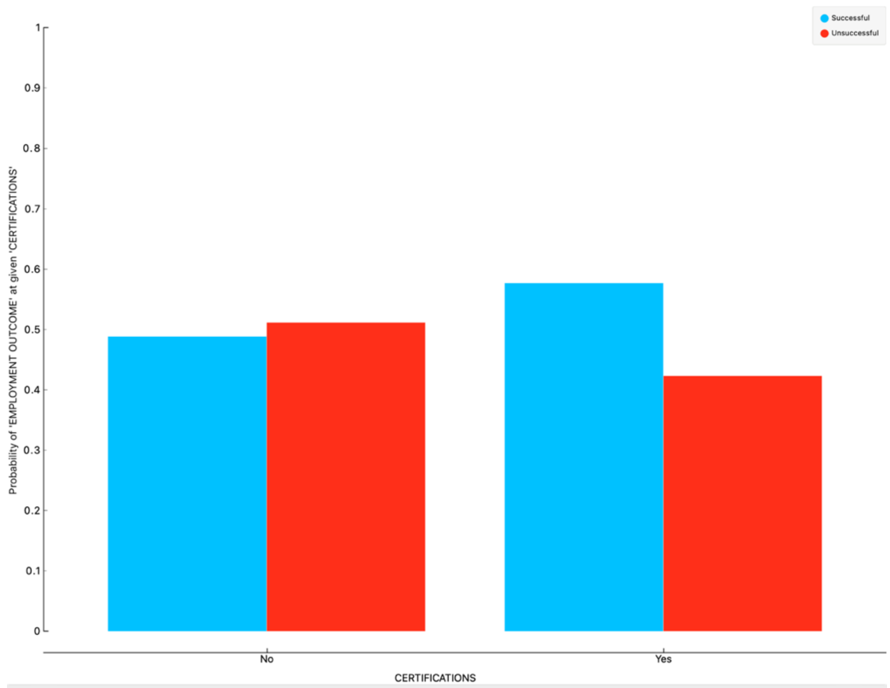

Figure 4.22 The probability of a successful (or unsuccessful outcome) given the possession of computing certifications

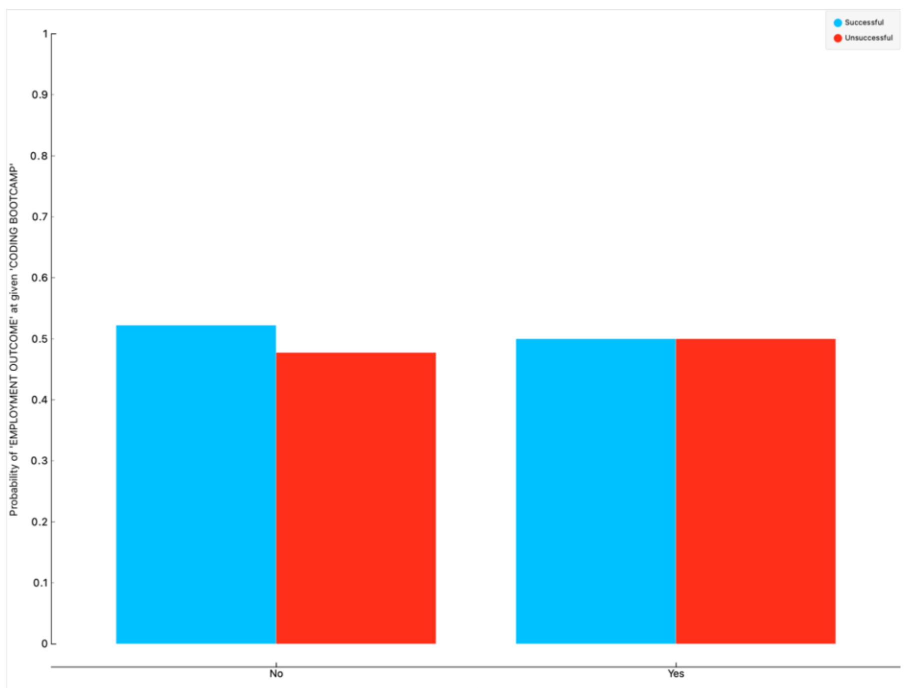

Figure 4.23 The probability of a successful (or unsuccessful outcome) given the attendance of a coding bootcamp

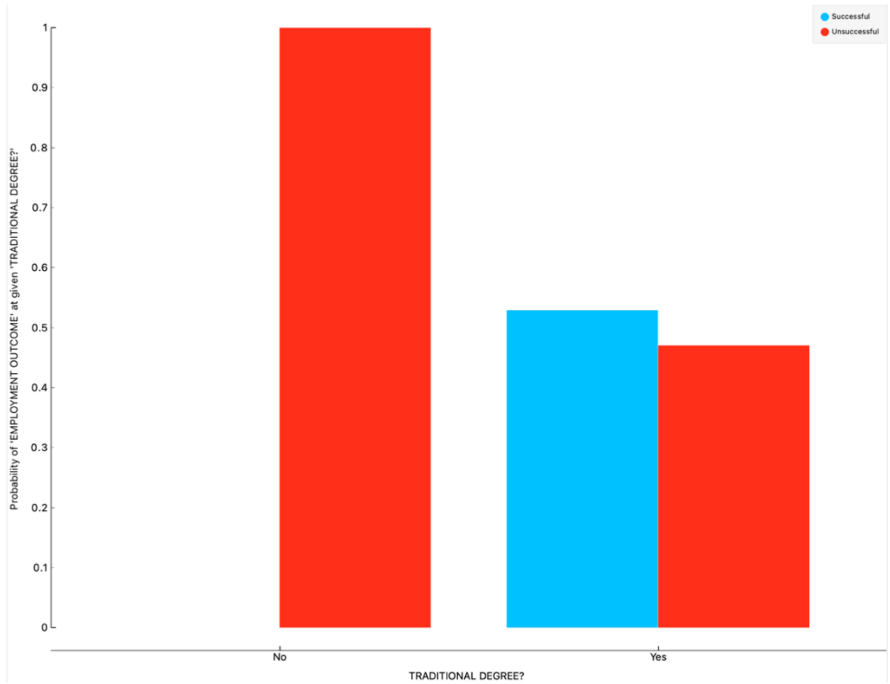

Figure 4.24 The probability of a successful (or unsuccessful outcome) given the possession of a traditional degree

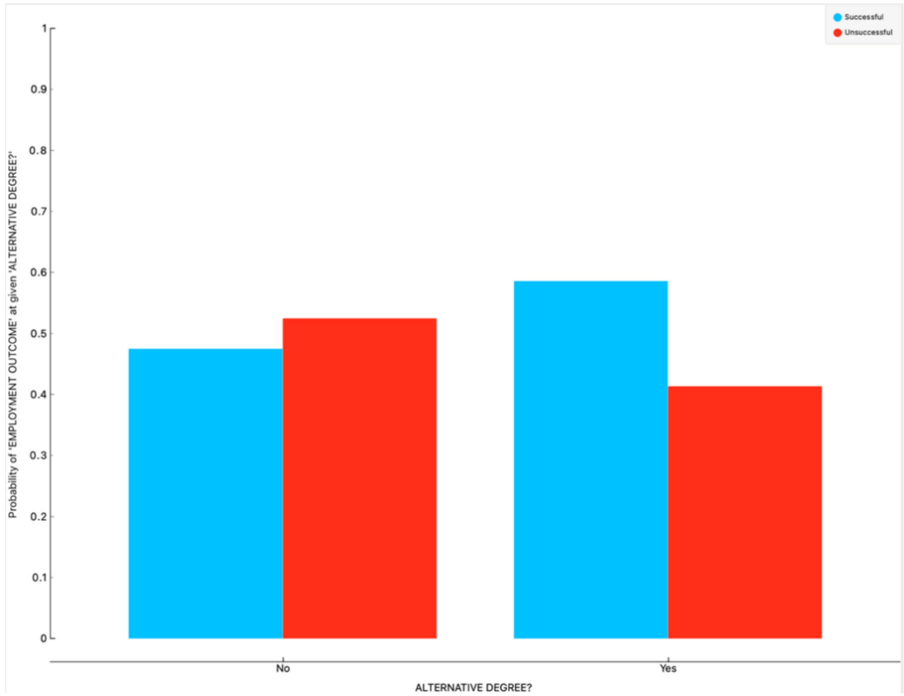

Figure 4.25 The probability of a successful (or unsuccessful outcome) given the possession of an alternative degree

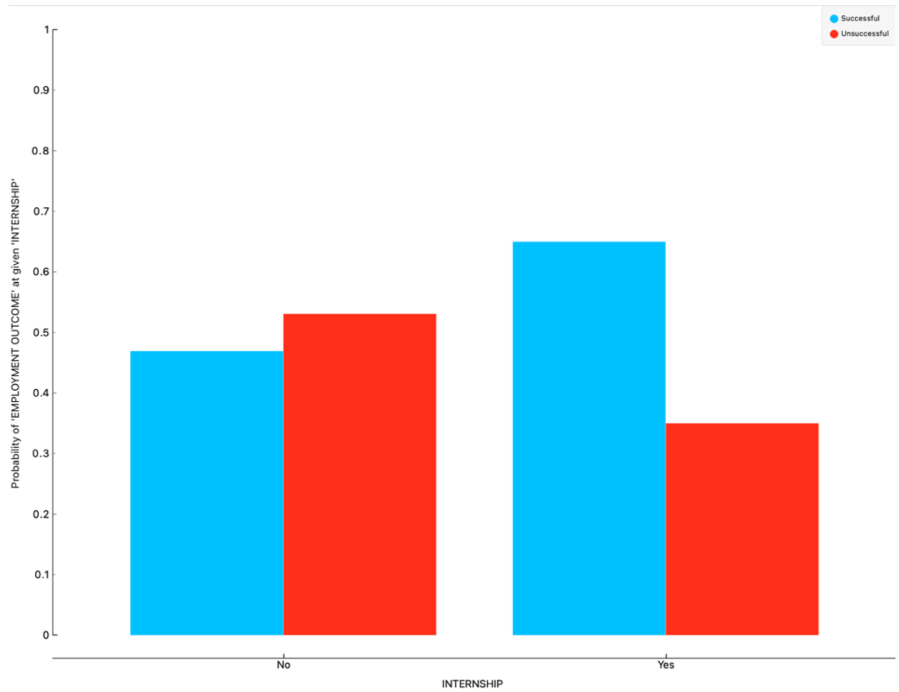

Figure 4.26 The probability of a successful (or unsuccessful outcome) given the possession of internship experience

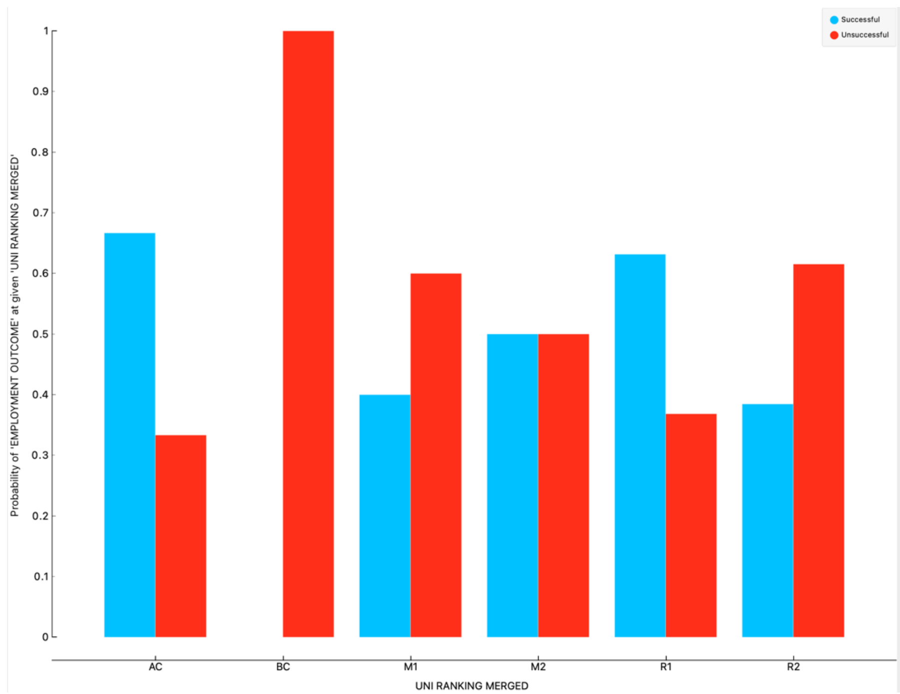

Figure 4.27 The probability of a successful (or unsuccessful outcome) given the ranking of their highest-ranked institution of study

Figure 4.22 shows that having a computing certification is a predictor of a successful employment outcome in the computing industry while a lack of a computing certification is a predictor of an unsuccessful employment outcome. In Figure 4.23, it is shown that not attending a coding bootcamp results in a higher chance of being successful than being unsuccessful, whereas attending a coding bootcamp produces an equal probability of successful and unsuccessful employment outcomes. The possession of a traditional degree (that is; associate's bachelor's, master's, or doctoral degree) is a predictor of a successful employment outcome in the computing industry while a lack of a traditional degree is a 100% predictor of an unsuccessful employment outcome. This can be seen in Figure 4.24. Similarly, Figure 4.25 shows that the possession of an alternative degree (that is; certifications or coding bootcamps) is a predictor of a successful employment outcome in the computing industry while a lack of an alternative degree is a predictor of an unsuccessful employment outcome.

The possession of internship experience also has a similar impact on the employment outcomes of women and blacks. Having a computing internship experience is a predictor of a successful employment outcome in the computing industry while a lack of computing internship experience is a predictor of an unsuccessful employment outcome, as seen in Figure 4.26. In Figure 4.27, the ranks of institutions that predict a successful employment outcome include: the AC (Associate-degree granting) and R1 (Very high research activity) institutions. On the other hand, the BC (Bachelor-granting only), M1 (Masters' – larger programs), and R2 (High research activity) institutions predict an unsuccessful employment outcome.

Some of these results are not generalizable, due to the limited amount of data on which the results are based. Within the data, only 3 respondents attended AC institutions and 2 were successful; only 1 respondent did not possess a traditional degree and the person had an unsuccessful outcome; only 2 respondents attended a coding bootcamp and one was successful while the other had an unsuccessful outcome; only 2 respondents had some college as their highest degree and both were successful; only 1 respondent had a high school or lower as their highest degree and the person had an unsuccessful outcome.

Similarly, using 2 clusters (k = 2) for this data analysis yields a silhouette value of 0.123 which is an average score. A silhouette value usually ranges from -1 to 1 where a score of 1 means that the data points are close to other data points within the same cluster and are far from other clusters. A score of -1, on the other hand, reveals the opposite; that is, the data points might likely be in the wrong cluster. Achieving an average silhouette score (0.123) in this result means that the clusters are very close together and that there might be outliers in the data.

The inability to generalize the results due to limited data, and the average silhouette scores due to outliers reveal that the results that are described above would need to be validated with larger and more refined datasets.

From the clustering results, the educational factors that are predictors of successful outcomes (when possessed by women and blacks) and whose absence results in unsuccessful employment outcomes for women and blacks are: Internship, Alternative degree, Traditional degree, Computing certifications, Computing degree, Masters' degree, Bachelor's degree, AC institution attendance, and R1 institution attendance. The impacts of these educational factors are measured individually, and not as a combination of factors as seen in section 4.1.3.7.

4.1.5 Classification Data Analysis Results

There are 5 measures that describe the results of the 4 classification models: Area under the curve (AUC), Classification Accuracy (CA), Precision, Recall, and F1.

AUC measures the ability of a classification model to distinguish between classes; the classes being "Successful" and "Unsuccessful" in this context. Models with higher AUCs are better at distinguishing classes than those with lower AUCs. AUC values below 0.7 indicate an inability to properly distinguish between classes while AUC values between 0.7 and 1 indicate an increasing ability to distinguish between classes.

Classification Accuracy (CA) measures the percentage of records that were correctly classified by the classification algorithm. That is, number of correct predictions divided by the total number of predictions. Values close to 1 indicate a high accuracy.

Precision measures the percentage or proportion of true positives. That is, the number of correct positive predictions divided by the total number of positive predictions. This is equivalent to $\textbf{number of true positives} / (\textbf{number of true positives} + \textbf{number of false positives})$. Values close to 1 indicate a high precision. For instance, a precision of 0.6 on the "Successful" target class means that a "Successful" prediction is correct only 60% of the time.

Recall measures the proportion of the number of true positives to the total number of actual positive outcomes. That is,

$$\textbf{number of true positives} \,/ (\textbf{number of true positives} \ + \ \textbf{number of false negatives}) \ .$$

Values close to 1 indicate a high recall.

The F1 score measures the weighted average of Precision and Recall.

(Arafiyah et al., 2018)

4.1.5.1 First Phase of Classification results

In the first phase of classification results, the full educational and employment histories of respondents are used as predictors for their employment outcome, and the performance of the prediction models are reported. The educational and employment variables include: Associate degree, Bachelor’s degree, Master’s degree, Doctoral degree, Highest degree attained, Certification, Coding bootcamp, Traditional degree, Alternative degree, Computing degree, Institution ranking, Internship, Current employment status, Time elapsed before employment, Persistence in computing, Highest employer ranking, Highest salary; where the last 5 variables are the employment variables. Table 4.2 shows the measures of the abilities of the 4 classification models to predict a “Successful” employment outcome given the full educational variables and employment variables.

Table 4.2 Classification results with full educational variables and employment variables

	AUC	CA	Precision	Recall	F1
Random Forest	0.838	0.733	0.704	0.863	0.775
Neural Network	0.839	0.760	0.724	0.887	0.798
Naïve Bayes	0.859	0.753	0.747	0.812	0.778
Logistic Regression	0.891	0.800	0.772	0.887	0.826

The AUC values show that Logistic Regression is the classification model with the best ability to distinguish between the “Successful” and “Unsuccessful” target classes. Using Logistic regression, 80% of the records were correctly classified. Of all the 4 classification models, Logistic Regression has the highest classification accuracy. Similar to the AUC and CA results, Precision, Recall, and F1 show that Logistic Regression possesses the highest precision and recall, of the 4 classification models under observation. This proves that the Logistic regression results in section 4.1.2 and 4.1.3 are reliable.

In the second part of the first phase, the merged educational variables and the employment variables are predictors for the employment outcome target variable. Then, the performance of the prediction models is reported. The educational and employment variables include: Traditional degree, Alternative degree, Computing degree, Institution ranking, Internship, Current employment status, Time elapsed before employment, Persistence in computing, Highest employer ranking, Highest salary; where Traditional degree includes Associate degree, Bachelor’s degree, Master's degree, and Doctoral degree; Alternative degree includes Certification and Coding bootcamp; and the last 5 variables are the employment variables. Table 4.3 shows the measures of the abilities of the 4 classification models to predict a "Successful" employment outcome given the merged educational variables and employment variables.

Table 4.3 Classification results with merged educational variables and employment variables

	AUC	CA	Precision	Recall	F1
Random Forest	0.860	0.760	0.712	0.925	0.804
Neural Network	0.884	0.807	0.780	0.887	0.830
Naïve Bayes	0.897	0.800	0.755	0.925	0.831
Logistic Regression	0.909	0.833	0.796	0.925	0.855

Similar to the results from the full educational variables and employment variables, the Logistic Regression classification model has the best performance of all the 4 classification models with a 90% AUC score, 83% classification accuracy, 79% precision, and 92% recall.

4.1.5.2 Second Phase of Classification results

In the first phase, the 5 employment variables were included in the list of predictors of the employment outcome target. Since the employment outcome variable was derived from those 5 employment variables, it seems reasonable that the prediction accuracy of the "Successful" employment outcome will be high when the 5 variables are included as part of the predictors. A high prediction accuracy of the "Successful" employment outcome proves that there is a high correlation between the 5 employment variables and the employment outcome variable.

In the second phase, the 5 employment variables are excluded from the list of predictors and the ability of the educational variables (alone) to predict the employment outcome variable is reported. The full list of educational variables include: Associate degree, Bachelor's degree, Master's degree, Doctoral degree, Highest degree attained, Certification, Coding bootcamp, Traditional degree, Alternative degree, Computing degree, Institution ranking, Internship. Table 4.4 shows the measures of the abilities of the 4 classification models to predict a "Successful" employment outcome given the full educational variables.

Table 4.4 Classification results with full educational variables

	AUC	**CA**	**Precision**	**Recall**	**F1**
Random Forest	0.557	0.533	0.549	0.700	0.615
Neural Network	0.634	0.620	0.607	0.812	0.695
Naïve Bayes	0.579	0.580	0.583	0.750	0.656
Logistic Regression	0.609	0.560	0.565	0.762	0.649

As seen by the AUC scores in the table above, the Neural Network model is the classification model with the best ability to distinguish between the "Successful" and "Unsuccessful" target classes. The Neural Network model also shows better performance with classification accuracy, precision, and recall. Despite the great performance of Neural Networks, it shows lesser performance than what is seen in the first phase of the classification analysis. This shows that educational history alone is not enough to accurately predict employment outcome.

4.1.6 Putting it all together

The data analysis results seen in the previous sections have answered the first two research questions by showing the following:

- According to the data, the typical educational pathway taken by women and blacks includes: Possession of a traditional degree, Attending a Rank1 institution, no alternative degree, and no internship. This can be seen in section 4.1.1
- Logistic Regression predicts the employment outcomes of blacks and women more accurately than Naïve Bayes, Neural Network, and Random Forest classification models. This can be seen in section 4.1.5.1.
- Using Logistic Regression and Predicted probabilities to study the relationships between the dependent variables and the target variable shows that the pathway that has the least probability of yielding a successful outcome for women and blacks is: Attending only a Rank 1 institution or having only a traditional degree. This is the pathway that the majority of women and black people have taken, as seen in the first bullet point above. This provides an explanation as to why women and blacks have not experienced a lot of successful computing employment outcomes. This can be seen in section 4.1.3.7.
- Using Logistic Regression and Predicted probabilities to study the relationships between the dependent variables and the target variable shows that the pathway that has the highest

probability of yielding a successful outcome for women and blacks is: Doing an internship, Possessing a traditional degree, Possessing an alternative degree, and Attending a Rank3 institution. Making all four educational choices yields over 90% probability of a successful employment outcome. This can be seen in section 4.1.3.7.

- The results of the Clustering analysis are similar to the Logistic regression and Predicted probabilities results. The difference is that the clustering results measures the impact of individual educational choices on the employment outcomes of blacks and women. From the clustering results, the educational factors that are predictors of successful outcomes (when possessed by women and blacks) and whose absence results in unsuccessful employment outcomes for women and blacks are: Internship, Alternative degree, Traditional degree, Computing certifications, Computing degree, Masters' degree, Bachelor's degree, AC institution attendance, and R1 institution attendance. In other words, doing an internship, possessing an alternative degree, possessing a traditional degree, attending a Rank3 institution, and attending a Rank1 institution all individually predict a successful employment outcome for women and blacks. This can be seen in section 4.1.4.
- It was shown that educational history alone is not enough to accurately predict the employment outcome of women and blacks as seen in section 4.1.5.2.
- Finally, more conclusive and generalizable results can be made if there is a larger dataset.

4.1.7 Discussion of Results

This research has shown that attending a Rank3 institution, that is, attending an associate-degree-granting college (e.g., community colleges) contributes to a successful employment outcome for women and blacks. This is in sharp contrast to the high probability of an unsuccessful employment outcome when women and blacks attend only a Rank1 institution, that is, an R1 or R2 institution (doctoral-degree-granting institutions with high or very high research activity).

There could be a number of reasons why this is the case. Research carried out by Sax et al. (2018) show that one of the reasons why women and people of color drop out of computing disciplines in college is because they do not have a sense of belonging in the field. On the other hand, a sense of belonging has been shown to be a predictor of success in college (Sax et al., 2018). How does this sense of belonging distinguish between Rank1 and Rank3 institutions being on the pathway to unsuccessful employment outcomes and successful outcomes, respectively?

Schwartz (2020) reports that first generation and underrepresented minority students (of which women and blacks are a part) who attend community colleges (a Rank3 institution) feel a stronger sense of belonging at these colleges, compared to their colleagues. On the other hand, non-first-generation students and students who do not identify as underrepresented minorities have a stronger sense of belonging at four-year colleges (of which Rank1 colleges are a type) than other students at their college.

It follows that a strong sense of belonging at community colleges would result in successful academic performance and successful employment outcomes. This can be seen in research carried out by Haberler and Levin (2013) which shows that student achievement at community colleges is improved in the presence of cohesion, connection, cooperation, and consistency.

For community college students who transfer to four-year colleges, transferring to a large four-year college has been shown to have a negative impact on their success (Umbach et al., 2018). However, transferring to a Historically Black College or University (HBCU) has a positive impact on student success and degree completion (Umbach et al., 2018). This might be because underrepresented minorities (blacks, in this case) continue to experience a sense of belonging while attending a four-year HBCU.

This discourse has shown how a sense of belonging within Rank3 institutions would propel underrepresented minorities along that pathway, ultimately leading to successful employment outcomes. There are many other factors that distinguish between the path that leads to successful employment outcomes versus the path that leads to unsuccessful employment outcomes for women and blacks. It might be helpful to consider factors like institution size or population, student-to-teacher ratio, teacher-as-mentor versus teacher-as-dictator model, whether the content taught is geared toward more theoretical knowledge or more practical knowledge, among other things. This is a possible area for future inquiry and research.

4.1.8 Application of these results to future BPC efforts

Given the data analysis results earlier shown, how can the extracted knowledge be of benefit to the broadening participation efforts in the computing field? There are two possible courses of action that will yield desirable results for BPC efforts: Recruiting more students to the most successful pathways and Removing barriers on the least successful pathways. This aims to answer the third research question.

4.1.8.1 Removing barriers on the least successful pathways

There are several reasons why the Rank 1 institution and traditional degree pathways do not result in successful outcomes. A reason could be the low representation of blacks and women on these pathways which do not provide sufficient data to paint an accurate picture of the effectiveness (or otherwise) of the traditional pathways in producing employment success. The barriers to increasing representation on these traditional pathways can be tackled and removed in order to attract more blacks and women to these pathways. Barriers such as the lack of a sense of belonging and the unconducive university learning environments for blacks and women (as highlighted in section 4.1.7 above and in section 2.3.2) need to be addressed. A re-evaluation can then take place to ascertain whether the unsuccessful outcomes are as a result of insufficient data or whether the traditional pathways simply do not work for women and blacks.

Since a barrier to the representation of women and blacks is their limited prior technical experience, modifying the curriculum on the traditional pathway to cater to the unique experiences of women and blacks by providing instruction suited to their technical expertise level could be an effort that would yield desirable results including a rise in representation of women and blacks, and even better employment outcomes after their education.

4.1.8.2 Recruiting more blacks and women to the most successful pathways

A more efficient approach to increasing the participation of blacks and women in computing employment would be to recruit them onto pathways that have been proven to result in successful employment outcomes for them. This dissertation research has shown that attending an associate-degree-granting institution, possessing a traditional degree, having an internship, and possessing an alternative degree are educational choices that make up the most successful pathways.

Therefore, efforts should be targeted towards recruiting more blacks and women into community colleges. Educational schemes and programs that enable blacks and women to attend a community college and then transition to a four-year university should also be invested in.

At the high school level, school counselors should be equipped to guide black and female students toward a computing program at a community college. The counselors should also be privy to the educational programs that provide a community college to university pathway for students, so that students are aware of this pathway into computing. For students who are unable to attend community college, information about coding bootcamps and certification can be provided to them.

For students who proceed to community college and the university (through the proposed hybrid program), the program directors or undergraduate coordinators should provide students with internship and certification opportunities so that by the end of their program, they would have travelled the educational pathway that results in successful computing employment opportunities.

CHAPTER V

CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

This research studied real-world data of working professionals, extracted from LinkedIn, in order to provide a picture of what educational choices and pathways have historically worked to provide successful computing employment outcomes for blacks and females, who are underrepresented in the field of computing. As seen in sections 2.1.3 and 3.3.2.2 of this dissertation, a person had a successful employment when they met all or 4 of the following 5 conditions: They were currently employed (as at the time of collecting the data used in this research), they had a moderate wait time (a year or less) before their first computing job, they persisted in a computing job (for at least 3 years), they were employed by a highly-ranked employer (a Fortune 500 company), and they earned a good annual income ($50,000 or higher).

After carrying out some descriptive statistics, logistic regression, predictive analytics, clustering, and classification on the pre-processed data, it was discovered that the majority of blacks and women did not do an internship (> 69%), had a traditional degree (> 93%), did not have an alternative degree (> 55%), and attended a Rank1 institution (R1, R2, and Doctoral/Professional universities) (> 66%). The data also showed that this popular pathway taken by women and blacks has not yielded the most successful computing employment outcomes for them. Rather, the most effective educational choices for successful computing employment outcomes are a Rank 3 institution education, possession of a traditional degree (associates degree, bachelor's degree, master's degree, doctoral degree), possession of an alternative degree (certifications, coding bootcamp degrees, or any other non-traditional degree), and undertaking a computing internship. This result is line with Blaney (2020)'s recommendation (seen in Section 2.4.2.4) that BPC efforts should be targeted towards the upward transfer students (students who transfer from community colleges to 4-year computing college programs) since there is a high percentage of underrepresented students among the upward transfers. This research has shown that, not only is there a high percentage of underrepresented students on the community college-to-4year college pathway, but this pathway also has the highest probability of resulting in successful employment outcomes for underrepresented minorities (women and blacks, in particular).

This dissertation set out to answer a vital question for the "Broadening Participation in Computing" (BPC) community: "What is a more effective strategy to increase the representation of women and blacks in the computing field and workforce?". This question has been answered by posing another question: "What if we broaden participation by identifying the successful pathways and recruiting underrepresented minorities onto them rather than focusing on the unsuccessful pathways?". This resulted in the question: "What then are the pathways that have resulted in the most successful employment outcomes for blacks and women?". The results of the data analysis within this research have provided an answer to the final question about the most successful pathways and has provided a solid foundation to be built upon as the BPC research community continue to answer the question about increasing the representation of blacks and women in the computing field and workforce.

5.2 FUTURE WORK

Given the results of this research and the conclusion above, there are a number of possible future research directions from this point.

First, the reasons why an education at a Rank 3 institution (that is, Associate-degree-granting institution) yields a successful employment outcome for women and blacks should be studied. For example, a closer look needs to be taken at the Rank 1 institutions (R1, R2, and Doctoral/Professional universities) and the Rank 3 institutions to compare and contrast the existing support services for black and female students within the institutions. This is important because lack of support for minorities has been identified as one of the factors that has contributed to the underrepresentation of these minorities in computing education. According to the results of this research, Rank 1 institutions appear to be lagging behind Rank 3 institutions in providing this support to underrepresented minorities. Therefore, comparing Rank 1 and Rank 3 institutions to determine what support services place Rank 3 institutions on the optimal educational pathway to successful computing employment outcomes for blacks and women is a recommended future research direction based on this work. Carrying out this comparison among Rank 1 institutions to determine how the more impactful Rank 1 institutions are able to serve the underrepresented minority leading to a successful employment outcome, is also a future research direction.

Similar to the first research direction above, an investigation into the reason why getting a traditional degree, in addition to attending a Rank 3 institution, is a predictor of successful computing employment outcomes for blacks and women is a future research direction.

Furthermore, the reason for the effectiveness of alternative degrees and internships is a viable subject for future investigation.

Second, developing and implementing strategies for recruitment of blacks and women onto the successful educational pathway (consisting of a Rank 3 institution education, traditional degree, alternative degree, and internship) should be researched further. The results of these efforts on the employment outcomes of blacks and women should then be evaluated. Also, the barriers on the least successful pathways should be identified and tackled, so that there can be a higher recruitment of females and blacks to those pathways. This will further increase the representation of blacks and women in the computing workforce.

With regard to data analysis techniques, a variety of data analysis techniques (more suitable to social media data) should be explored. Artificial Intelligence techniques that infuse fuzzy logic into their operations (such as the Neural Fuzzy Networks) can also be explored. This would require a larger dataset than what was employed within this research. Therefore, future research should involve larger datasets that are more representative of the general population and datasets that are able to produce analysis results that are generalizable to the entire population.

REFERENCES

Adrion, W. R., Biskup, S., Boisvert, D., Clarke, L., Fountain, J., Grocer, P., ... & Wiens, A. (2008). Broadening participation in computing: K12-community-college-university-graduate pathways. In *2008 38th Annual Frontiers in Education Conference* (pp. S4F-15). IEEE.

Aish, N., Asare, P., & Miskioğlu, E. E. (2017). People like me increasing likelihood of success for underrepresented minorities in stem by providing realistic and relatable role models. In *2017 IEEE Frontiers in Education Conference (FIE)* (pp. 1-4). IEEE.

Aish, N., Asare, P., & Miskioğlu, E. E. (2018, March). People like me: Providing relatable and realistic role models for underrepresented minorities in STEM to increase their motivation and likelihood of success. In *2018 IEEE integrated STEM education conference (ISEC)* (pp. 83-89). IEEE.

Ajilore O. (2020). On the Persistence of the Black-White Unemployment Gap. Retrieved from: https://www.americanprogress.org/issues/economy/reports/2020/02/24/480743/persistence-black-white-unemployment-gap/

Allison, P. (2012). When Can You Safely Ignore Multicollinearity? (Statistical Horizons). Retrieved from: https://statisticalhorizons.com/multicollinearity

Alvarado, C., Dodds, Z., & Libeskind-Hadas, R. (2012). Increasing women's participation in computing at Harvey Mudd College. *ACM Inroads, 3*(4), 55-64.

Alvarado, C., & Judson, E. (2014). Using targeted conferences to recruit women into computer science. *Communications of the ACM, 57*(3), 70-77.

Arafiyah, R., Hermin, F., Kartika, I. R., Alimuddin, A., & Saraswati, I. (2018, November). Classification of Dengue Haemorrhagic Fever (DHF) using SVM, naive bayes and random forest. In *IOP Conference Series: Materials Science and Engineering* (Vol. 434, No. 1, p. 012070). IOP Publishing.

Bannikova, L. N., Boronina, L. N., & Baliasov, A. A. (2018). Industry-University Collaboration in Training Engineers (The Case of Sverdlovsk Region). 2018 IEEE International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS), Quality Management, Transport and Information Security, Information Technologies, (IT&QM&IS), 2018 IEEE International Conference, 824–827. https://doi.org/10.1109/ITMQIS.2018.8524957

Blackburn, R. D., Porto, S. C. S., & Thompson, J. J. (2016). Competency-based education and the relationship to digital badges. Taylor and Francis Inc. 30-38. https://doi.org/10.4324/9781315718569

Blaney, J. M. (2020, February). Broadening Participation in Computing: The Role of Upward Transfer. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (pp. 254-260).

Bowman, D. D. (2018). Declining Talent in Computer Related Careers. Journal of Academic Administration in Higher Education, 14(1), 1–4.

Brooke, V., & Revell, G. Quality Indicators for Competitive Employment Outcomes & Indicator Form Fact Sheet.

Brown, Q. (2016). BROADENING PARTICIPATION. *ACM Inroads*, *7*(4), 46-48.

Bruckman, A., Biggers, M., Ericson, B., McKlin, T., Dimond, J., DiSalvo, B., ... & Yardi, S. (2009). " Georgia computes!" improving the computing education pipeline. *ACM SIGCSE Bulletin*, *41*(1), 86-90.

Cambridge Dictionary (2021). Representation. https://dictionary.cambridge.org/us/dictionary/english/representation

Carreon, C. (2019). High school students will have faster pathway to computer science jobs through accelerated degree program. *Northeast Mississippi Daily Journal (Tupelo, MS).*

Castro, R., & Coen-Pirani, D. (2016). Explaining the evolution of educational attainment in the United States. *American Economic Journal: Macroeconomics*, *8*(3), 77-112.

Chookittikul, W., & Maher, P. E. (2011). Effective real-world project collaboration: Strategies from a cyber security degree program. 2011 24th IEEE-CS Conference on Software Engineering Education and Training (CSEE&T), Software Engineering Education and Training (CSEE&T), 2011 24th IEEE-CS Conference On, 429–433. https://doi.org/10.1109/CSEET.2011.5876119

Clear, A., Parrish, A. S., Impagliazzo, J., & Zhang, M. (2019, February). Computing Curricula 2020: introduction and community engagement. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 653-654).

Commission on Higher Education & Employability (2018). Learning for Life and Work. New England Board of Higher Education. https://nebhe.org/info/pdf/policy/Learning_for_Life_and_Work_Report.pdf

ComputerScience Staff (2021). Top Careers in Computer Science. Retrieved from: https://www.computerscience.org/careers/. Accessed: 21st July 2021.

Dahlberg, T., Barnes, T., Buch, K., & Rorrer, A. (2011). The STARS alliance: Viable strategies for broadening participation in computing. *ACM Transactions on Computing Education (TOCE)*, *11*(3), 1-25.

Dasgupta, N., & Stout, J. G. (2014). Girls and women in science, technology, engineering, and mathematics: STEMing the tide and broadening participation in STEM careers. *Policy Insights from the Behavioral and Brain Sciences, 1*(1), 21-29.

Denning, P. J. (2007). The computing field: Structure. *Wiley encyclopedia of computer science and engineering*, 615-623.

Doerschuk, P., Liu, J., & Mann, J. (2009). Increasing participation of females and underrepresented minorities in computing. *Computer*, *42*(4), 110-113.

Dragoumanos, S., Kakarountas, A., & Fourou, T. (2017). Young technology entrepreneurship enhancement based on an alternative approach of project-based learning. 2017 IEEE Global Engineering Education Conference (EDUCON), Global Engineering Education Conference (EDUCON), 2017 IEEE, 351–358. https://doi.org/10.1109/EDUCON.2017.7942872

DuBow, W. & Gonzalez, J.J. (2020). NCWIT Scorecard: The Status of Women in Technology. Boulder, CO: NCWIT.

DuBow, W. M., Quinn, B. A., Townsend, G. C., Robinson, R., & Barr, V. (2016). Efforts to make computer science more inclusive of women. *ACM Inroads, 7*(4), 74-80.

El-Agamy, R., & Tsuda, K. (2013). Analysis of the Job Categories of the New Japanese Information Technology Skills Standards. COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE, 246, 268.

Eurofound (2016). The gender employment gap – challenges and solutions. Retrieved from: https://www.eurofound.europa.eu/the-gender-employment-gap-challenges-and-solutions

Evans, T. (2007). “Skills gap” remains a drag on economy: State’s growth maintains pace. Daily Oklahoman, The (OK)

Exter, M., Ashby, I., & Shaurette, M. (2015, June). Entering the first year of a multidisciplinary, hands-on, competency-based learning experience: Hopes and concerns of students, parents, and faculty. In *2015 ASEE Annual Conference & Exposition* (pp. 26-677).

Exter, M., Caskurlu, S., & Fernandez, T. (2018). Comparing Computing Professionals’ Perceptions of Importance of Skills and Knowledge on the Job and Coverage in Undergraduate Experiences. ACM TRANSACTIONS ON COMPUTING EDUCATION, 18(4). https://doi.org/10.1145/3218430

Fealing, K. H., Lai, Y., & Myers Jr, S. L. (2015). Pathways vs. pipelines to broadening participation in the STEM workforce. *Journal of Women and Minorities in Science and Engineering, 21*(4).

Freeman, J., Magerko, B., Edwards, D., Moore, R., McKlin, T., & Xambó, A. (2015). EarSketch: A STEAM approach to broadening participation in computer science principles. In *2015 Research in Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT)* (pp. 1-2). IEEE.

Funk, C., & Parker, K. (2018). Women and men in STEM often at odds over workplace equity.

Gawlowicz, S. (2007). Policymakers urged to address concerns about US science and technology workforce (Rochester Institute of Technology). Retrieved from http://www.rit.edu/news/story.php?id=45854

Glassdoor (2021). Glassdoor Salaries. Retrieved from: https://www.glassdoor.com/Salaries/index.htm

Google (2021). Standard error. Retrieved from: https://www.google.com/search?client=firefox-b-d&q=standard+error

Google (Firm) Gallup (Firm). (2016). Diversity gaps in computer science: exploring the underrepresentation of girls, Blacks and Hispanics.

Haberler, Z., & Levin, J. S. (2014). The four Cs of promising practices in community colleges. *Community College Journal of Research and Practice, 38*(5), 403-416.

Hoffman, B., Morelli, R., & Rosato, J. (2019). Student engagement is key to broadening participation in CS. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 1123-1129).

Hossain (2019). What is Univariate, Bivariate and Multivariate analysis? (HotCubator). Retrieved from: https://hotcubator.com.au/research/what-is-univariate-bivariate-and-multivariate-analysis/

Indeed (2021). Indeed Salaries. Retrieved from: https://www.indeed.com/career/salaries

Indeed Editorial Team (2021). *How Long Does It Take To Find a Job?* Indeed. Retrieved from: https://www.indeed.com/career-advice/finding-a-job/how-long-does-it-take-to-find-a-job

Indeed Editorial Team (2021). *21 Different Types of IT Jobs to Explore* Indeed. Retrieved from: https://www.indeed.com/career-advice/finding-a-job/types-of-it-jobs

Jacobs, H. (2015). Teaching for Employability: Key Concepts and Best Practice Principles. International Conference on Education and New Developments (END 2015), 91-95. https://files.eric.ed.gov/fulltext/ED578161.pdf

Job Categories and Education Requirements in Information Technology. (2000). Techniques: Connecting Education and Careers, 75(7), 28–31.

Joyce. (n.d.). *Advantages & Disadvantages of Working for a Fortune 500 Employer*. Job-hunt. Retrieved from: https://www.job-hunt.org/employers/Fortune-500/why-fortune-500-jobs.shtml

Kannankutty, N. (2007). Multiple Pathways to Science and Engineering Employment: Characteristics of the U.S. S&E Workforce. 2007 Atlanta Conference on Science, Technology and Innovation Policy, Science, Technology and Innovation Policy, 2007 Atlanta Conference On, 1–8. https://doi.org/10.1109/ACSTIP.2007.4472909

Kromm, C. (2011). The south's economic crisis. Retrieved from: https://www.facingsouth.org/2011/10/the-souths-economic-crisis.html

Learn How To Become (n.d.). Computer Science Degrees & Careers: How to Work in Computer Science. Retrieved from: https://www.learnhowtobecome.org/computer-careers/.

Lee, S. B., Sun, R. A., & Lynn, R. (2019). Mississippi coding academies: A nontraditional approach to computing education. *ASEE Annual Conference and Exposition, Conference Proceedings.*

Leibert, A. (2014). *Measuring Employment Outcomes for Graduates*. Minnesota Department of Employment and Economic Development. https://mn.gov/deed/newscenter/publications/trends/march-2014/employment-outcome-graduate.jsp.

Lindsay, J., & Babb, S. (2015). Measuring Employment Outcomes for Workforce Development. USA; FHI 360.

LinkedIn (2020). Andela. Available at: https://www.linkedin.com/company/andela

Liu, H., & Cocea, M. (2017). Semi-random partitioning of data into training and test sets in granular computing context. *Granular Computing*, *2*(4), 357-386.

Llorens, A., Chiaramonte, L., Llinàs-Audet, X., & Ras, A. (2013). The ICT skills gap in Spain: Industry expectations versus university preparation. Computer Applications in Engineering Education, 21(2), 256–264. https://doi.org/10.1002/cae.20467

Locsin, A. (n.d.). Careers in the Computer field (Chron). Retrieved from: https://work.chron.com/careers-computer-field-4158.html

Lynn, K. M., Raphael, C., Olefsky, K., & Bachen, C. M. (2003). Bridging the gender gap in computing: An integrative approach to content design for girls. *Journal of educational computing research*, *28*(2), 143-162.

Margolis, J., Ryoo, J. J., Sandoval, C. D., Lee, C., Goode, J., & Chapman, G. (2012). Beyond access: Broadening participation in high school computer science. *ACM Inroads*, *3*(4), 72-78.

McClelland, M. (2001). Closing the IT gap for race and gender. *Journal of Educational Computing Research*, *25*(1), 5-15.

McCrone, T., White, R., Kettlewell, K., Sims, D., Rush, C., & National Foundation for Educational Research (NFER) (United Kingdom). (2019). Evaluation of University Technical Colleges. In National Foundation for Educational Research. National Foundation for Educational Research.

MDRC (2015). *New Pathways to Careers and College – Examples, Evidence, and Prospects*

Merriam-Webster. (n.d.). Underrepresented. In *Merriam-Webster.com dictionary*. Retrieved May 10, 2021, from https://www.merriam-webster.com/dictionary/underrepresented

Merriam-Webster. (n.d.). Workforce. In *Merriam-Webster.com dictionary*. Retrieved May 10, 2021, from https://www.merriam-webster.com/dictionary/workforce

Miller, K. W., & Voas, J. (2008). Computer Scientist, Software Engineer, or IT Professional: Which Do You Think You Are? IT Professional, IT Prof, 10(4), 4–6. https://doi.org/10.1109/MITP.2008.64

Montandon, J. E., Politowski, C., Silva, L. L., Valente, M. T., Petrillo, F., & Guéhéneuc, Y. G. (2021). What skills do IT companies look for in new developers? a study with Stack Overflow Jobs. *Information and Software Technology*, *129*, 106429.

National Center for Education Statistics (NCES) [Online]. Available at: https://nces.ed.gov/

National Science Foundation (U.S.) Hagerstown Community College (2011). *Pathways to Cybersecurity and Information Assurance Careers*. Available at: http://www.nsf.gov/awardsearch/showAward?AWD_ID=1104195

National Science Foundation (2021). Broadening Participation in Computing (BPC). Retrieved from: https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505891

Negoita, M., & Dunham, K. (2013). Designing a Performance Measurement System for Career Pathways.

New York State Education Department (2018). Employment Outcome Policy. Retrieved from: http://www.acces.nysed.gov/vr/01000-employment-outcome-policy#:~:text=An%20employment%20outcome%20means%20an,%2C%20advancing%20in%2C%20or%20retaining%3A&text=Other%20types%20of%20employment%20in,employment%2Fbusiness%20ownership%20and%20telecommuting.

Nunn, S. (2019). The South's Economy is Falling behind: "All of a Sudden the Money Stops Flowing". Retrieved from: https://www.wsj.com/articles/the-souths-economy-is-falling-behind-all-of-a-sudden-the-money-stops-flowing-11560101610

Obama, B. (2007). Obama initiatives would make America more competitive by investing in diversity [Press release]. Retrieved from https://votesmart.org/public-statement/255254/obama-initiatives-would-make-america-more-competitive-by-investing-in-diversityindex#.Wq9aJahubcc

Panko, R. R. (2008). IT employment prospects: beyond the dotcom bubble. *European Journal of Information Systems, 17*(3), 182-197.

Peckham, J., Harlow, L. L., Stuart, D. A., Silver, B., Mederer, H., & Stephenson, P. D. (2007). Broadening participation in computing: issues and challenges. *ACM SIGCSE Bulletin*, *39*(3), 9-13.

Penn State (n.d.). Definitions. Retrieved May 10, 2021, from https://agsci.psu.edu/diversity/awareness/definitions#:~:text=Underrepresented%20Minority%20can%20be%20defined,the%20population%20in%20the%20country

Riel, A., Tichkiewitch, S., Stolfa, J., Stolfa, S., Kreiner, C., Messnarz, R., & Rodic, M. (2016). Industry-academia Cooperation to Empower Automotive Engineering Designers. Procedia CIRP, 50, 739–744. https://doi.org/10.1016/j.procir.2016.04.188

Sax, L. J., Blaney, J. M., Lehman, K. J., Rodriguez, S. L., George, K. L., & Zavala, C. (2018). Sense of belonging in computing: The role of introductory courses for women and underrepresented minority students. *Social Sciences*, *7*(8), 122.

Shackelford, R. L., McGettrick, A. D., Sloan, R. H., Topi, H., Davies, G., Kamali, R., Cross, J., Impagliazzo, J., LeBlanc, R. J., & Lunt, B. M. (2006). Computing Curricula 2005: The Overview Report. *Proceedings of the 37th SIGCSE Technical Symposium on Computer Science Education*, 456. https://doi.org/10.1145/1121341.1121482

Smith, M. (2016). *Computer Science For All*. National Archives and Records Administration. Retrieved from: https://obamawhitehouse.archives.gov/blog/2016/01/30/computer-science-all.

Someka Excel Solutions (2020). Fortune 500 Excel List. Retrieved from: https://www.someka.net/excel-template/fortune-500-excel-list/

StatisticsSolutions (2021). Assumptions of Logistic Regression. Retrieved from: https://www.statisticssolutions.com/assumptions-of-logistic-regression/#:~:text=First%2C%20logistic%20regression%20does%20not,Third%2C%20homoscedasticity%20is%20not%20required.

Stoops, N. (2004). Educational attainment in the United States: 2003. *Current population*.

Stout, J., & Camp, T. (2014). Now what? Action items from social science research to bridge the gender gap in computing research. *ACM SIGCAS Computers and Society*, *44*(4), 5-8.

Schwartz (2020). Minority and first-gen students feel greater belonging at 2-year colleges, study finds. Retrieved from: https://www.highereddive.com/news/minority-and-first-gen-students-feel-greater-belonging-at-2-year-colleges/569707/. Accessed: 30th September 2021.

TechRepublic (2016). If there's a tech skills shortage, why are so many computer graduates unemployed? Retrieved from: https://www.techrepublic.com/article/if-theres-a-tech-skills-shortage-why-are-so-many-computer-graduates-unemployed/. Accessed: 3rd Dec. 2019.

Trauth, E. M., Farwell, D. W., & Lee, D. (1993). The IS Expectation Gap: Industry Expectations Versus Academic Preparation. MIS Quarterly, 17(3), 293–307. https://doi.org/10.2307/249773

UCLA Institute for Digital Research & Education (n.d.). Logistic Regression Analysis | Stata Annotated Output. Retrieved from: https://stats.idre.ucla.edu/stata/output/logistic-regression-analysis/. Accessed: 20th June 2021.

Umbach, P. D., Tuchmayer, J. B., Clayton, A. B., & Smith, K. N. (2019). Transfer student success: Exploring community college, university, and individual predictors. *Community College Journal of Research and Practice*, *43*(9), 599-617.

University of Washington (2021). *What can students do to improve their chances of finding employment after college?* Disabilities, Opportunities, Internetworking, and Technology. https://www.washington.edu/doit/what-can-students-do-improve-their-chances-finding-employment-after-college#block-menu-block-3

U.S. Bureau of Labor Statistics (2021). Occupational Outlook Handbook: Computer and Information Technology Occupations. Retrieved from: https://www.bls.gov/ooh/computer-and-information-technology/home.htm. Accessed: 21st July 2021.

U.S. News (2021). Best Technology Jobs. Retrieved from: https://money.usnews.com/careers/best-jobs/rankings/best-technology-jobs. Accessed: 21st July 2021.

Van Epps, A. S., Ashby, I., Gray, C. M., & Exter, M. (2016). Supporting student attainment and management of competencies in a trans-disciplinary degree program. ASEE Annual Conference and Exposition, Conference Proceedings, 2016–June.

Varma, R. (2018). US science and engineering workforce: Underrepresentation of women and minorities. *American Behavioral Scientist*, *62*(5), 692-697.

Waguespack, L., Topi, H., Frezza, S., Babb, J., Marshall, L., Takada, S., Veer, G., & Pears, A. (2019). Adopting Competency Mindful of Professionalism in Baccalaureate Computing Curricula. 2019 Proceedings of the EDSIG Conference, 1-17.

Waters, J. K. (2008). A Pathway to Achievement. *T H E Journal*, 35(9), 32.

Webb, D. C., Repenning, A., & Koh, K. H. (2012, February). Toward an emergent theory of broadening participation in computer science education. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education* (pp. 173-178).

Weller, C.E. (2019). African Americans face Systematic Obstacles to Getting Good Jobs. Retrieved from: https://www.americanprogress.org/issues/economy/reports/2019/12/05/478150/african-americans-face-systematic-obstacles-getting-good-jobs/

Whitney, T., & Taylor, V. (2018). Increasing women and underrepresented minorities in computing: The landscape and what you can do. *Computer*, *51*(10), 24-31.

Wikipedia (2019). Alternative pathways in education. Retrieved from: https://en.wikipedia.org/wiki/Alternative_pathways_in_education

Wikipedia (2020). Andela. Retrieved from: https://en.wikipedia.org/wiki/Andela

Wikipedia (2021). LinkedIn. Retrieved from: https://en.wikipedia.org/wiki/LinkedIn

Wikipedia (2021). Orange (software). Retrieved from: https://en.wikipedia.org/wiki/Orange_(software)

Wikipedia (2021). Stata. Retrieved from: https://en.wikipedia.org/wiki/Stata

Williams J. & Wilson V. (2019). Black workers endure persistent racial disparities in employment outcomes. Retrieved from: https://www.epi.org/publication/labor-day-2019-racial-disparities-in-employment/

Women Driving Change: A Pathway to a Better Mississippi, Mississippi Black Women's Roundtable (MS BWR) and National Women's Law Center, 2019 joint report, Available: https://msblackwomensroundtable.org/wp-content/uploads/2019/09/final_nwlc_MS_Report.pdf

Workforce Connections (2015). *Measuring Employment Outcomes* [PowerPoint slides]. https://youtheconomicopportunities.org/sites/default/files/uploads/resource/Outcome_Measurement_Event_Presentation_Feb_5_2015.pdf

ZipRecruiter (2021). ZipRecruiter Salaries. Retrieved from: https://www.ziprecruiter.com/Salaries

Zweben, S., & Bizot, B. (2016). 2015 Taulbee survey. Computing Research News: CRA.

Zweben, S., & Bizot, B. (2017). 2016 Taulbee Survey. Computing Research News, 29(5), 3-49.

Zweben, S., & Bizot, B. (2018). 2017 CRA Taulbee Survey. Computing Research News, 30(5), 1-47.

Zweben, S., & Bizot, B. (2019). 2018 Taulbee Survey. Computing Research News, 31(5), 1-72.

APPENDIX A

IRB APPROVAL LETTER

MISSISSIPPI STATE UNIVERSITY™

Office of Research Compliance

Institutional Review Board for the Protection of
Human Subjects in Research
P.O. Box 6223
53 Morgan Avenue
Mississippi State, MS 39762
P. 662.325.3294

www.orc.msstate.edu

NOTICE OF DETERMINATION FROM THE HUMAN RESEARCH PROTECTION PROGRAM

DATE:	December 15, 2020
TO:	Shahram Rahimi, PhD, Computer Science and Engineering, John Swan;Michael Taquino John Swan, PhD, Computer Science and Engineering, Michael Taquino, PhD, nSPARC, Mercy Jaiyeola, MS, Computer Science & Engineering
PROTOCOL TITLE:	Identifying Pathways into the Computing Field and Workforce
FUNDING SOURCE:	
PROTOCOL NUMBER:	IRB-20-479
	Approval Date: December 15, 2020 Expiration Date: December 14, 2025

EXEMPTION DETERMINATION

The review of your research study referenced above has been completed. The HRPP had made an Exemption Determination as defined by 45 CFR 46.101(b)4. Based on this determination, and in accordance with Federal Regulations, your research does not require further oversight by the HRPP.

Employing best practices for Exempt studies is strongly encouraged such as adherence to the ethical principles articulated in the Belmont Report, found at www.hhs.gov/ohrp/regulations-and-policy/belmont-report/# as well as the MSU HRPP Operations Manual, found at www.orc.msstate.edu/humansubjects. As part of best practices in research, it is the responsibility of the Principal Investigator to ensure that personnel added after this Exemption Determination notice have completed IRB training prior to their involvement in the research study. Additionally, to protect the confidentiality of research participants, we encourage you to destroy private information which can be linked to the identities of individuals as soon as it is reasonable to do so.

Based on this determination, this study has been inactivated in our system. This means that recruitment, enrollment, data collection, and/or data analysis **CAN** continue, yet personnel and procedural amendments to this study are no longer required. **If at any point, however, the risk to participants increases, you must contact the HRPP immediately. If you are unsure if your proposed change would increase the risk, please call the HRPP office and they can guide you.**

If this research is for a thesis or dissertation, this notification is your official documentation that the HRPP has made this determination.

If you have any questions relating to the protection of human research participants, please contact the HRPP Office at irb@research.msstate.edu. We wish you success in carrying out your research project.

Review Type:	EXEMPT
IRB Number:	IORG0000467

Figure A.1 IRB Approval Letter